PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/12, C07K 14/47, A61K 38/17

A2

(11) International Publication Number: WO 98/46757

(43) International Publication Date: 22 October 1998 (22.10.98)

(21) International Application Number: PCT/US98/07999

(22) International Filing Date: 14 April 1998 (14.04.98)

(30) Priority Data:

08/843,374	15 April 1997 (15.04.97)	US
09/059,487	13 April 1998 (13.04.98)	US

(71) Applicant: GENETICS INSTITUTE, INC. [US/US]; 87 CambridgePark Drive, Cambridge, MA 02140 (US).

(72) Inventors: JACOBS, Kenneth; 151 Beaumont Avenue, Newton, MA 02160 (US). MCCOY, John, M.; 56 Howard Street, Reading, MA 01867 (US). LAVALLIE, Edward, R.; 113 Ann Lee Road, Harvard, MA 01451 (US). RACIE, Lisa, A.; 124 School Street, Acton, MA 01720 (US). MERBERG, David; 2 Orchard Drive, Acton, MA 01720 (US). TREACY, Maurice; 93 Walcott Road, Chestnut Hill, MA 02167 (US). SPAULDING, Vikki; 11 Meadowbank Road, Billeria, MA 01821 (US). AGOSTINO, Michael, J.; 26 Wolcott Avenue, Andover, MA 01810 (US).

(74) Agent: SPRUNGER, Suzanne, A.; Genetics Institute, Inc., 87 CambridgePark Drive, Cambridge, MA 02140 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published

Without international search report and to be republished upon receipt of that report.

(54) Title: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM

(57) Abstract

Polynucleotides and the proteins encoded thereby are disclosed.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM

This application is a continuation-in-part of application Ser. No. 60/XXX,XXX (converted to a provisional application from non-provisional application Ser. No. 08/843,374), filed April 15, 1997, which is incorporated by reference herein.

FIELD OF THE INVENTION

The present invention provides novel polynucleotides and proteins encoded by such polynucleotides, along with therapeutic, diagnostic and research utilities for these polynucleotides and proteins.

BACKGROUND OF THE INVENTION

Technology aimed at the discovery of protein factors (including e.g., cytokines, such as lymphokines, interferons, CSFs and interleukins) has matured rapidly over the past decade. The now routine hybridization cloning and expression cloning techniques clone novel polynucleotides "directly" in the sense that they rely on information directly related to the discovered protein (i.e., partial DNA/amino acid sequence of the protein in the case of hybridization cloning; activity of the protein in the case of expression cloning). More recent "indirect" cloning techniques such as signal sequence cloning, which isolates DNA sequences based on the presence of a now well-recognized secretory leader sequence motif, as well as various PCR-based or low stringency hybridization cloning techniques, have advanced the state of the art by making available large numbers of DNA/amino acid sequences for proteins that are known to have biological activity by virtue of their secreted nature in the case of leader sequence cloning, or by virtue of the cell or tissue source in the case of PCR-based techniques. It is to these proteins and the polynucleotides encoding them that the present invention is directed.

SUMMARY OF THE INVENTION

In one embodiment, the present invention provides a composition comprising an isolated polynucleotide selected from the group consisting of:

(a) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:1;

(b) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:1 from nucleotide 1799 to nucleotide 2332;

(c) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:1 from nucleotide 2288 to nucleotide 2332;

(d) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:1 from nucleotide 2306 to nucleotide 2754;

(e) a polynucleotide comprising the nucleotide sequence of the full-length protein coding sequence of clone en539_8 deposited under accession number ATCC 98408;

(f) a polynucleotide encoding the full-length protein encoded by the cDNA insert of clone en539_8 deposited under accession number ATCC 98408;

(g) a polynucleotide comprising the nucleotide sequence of a mature protein coding sequence of clone en539_8 deposited under accession number ATCC 98408;

(h) a polynucleotide encoding a mature protein encoded by the cDNA insert of clone en539_8 deposited under accession number ATCC 98408;

(i) a polynucleotide encoding a protein comprising the amino acid sequence of SEQ ID NO:2;

(j) a polynucleotide encoding a protein comprising a fragment of the amino acid sequence of SEQ ID NO:2 having biological activity, the fragment comprising the amino acid sequence from amino acid 84 to amino acid 93 of SEQ ID NO:2;

(k) a polynucleotide which is an allelic variant of a polynucleotide of (a)-(h) above;

(l) a polynucleotide which encodes a species homologue of the protein of (i) or (j) above ; and

(m) a polynucleotide that hybridizes under stringent conditions to any one of the polynucleotides specified in (a)-(j).

Preferably, such polynucleotide comprises the nucleotide sequence of SEQ ID NO:1 from nucleotide 1799 to nucleotide 2332; the nucleotide sequence of SEQ ID NO:1 from nucleotide 2288 to nucleotide 2332; the nucleotide sequence of SEQ ID NO:1 from nucleotide 2306 to nucleotide 2754; the nucleotide sequence of the full-length protein coding sequence of clone en539_8 deposited under accession number ATCC 98408; or the nucleotide sequence of a mature protein coding sequence of clone en539_8 deposited under accession number ATCC 98408. In other preferred embodiments, the polynucleotide encodes the full-length or a mature protein encoded by the cDNA insert of clone en539_8 deposited under accession number ATCC 98408. In yet other preferred embodiments, the present invention provides a polynucleotide encoding a protein comprising the amino acid sequence of SEQ ID NO:2 from amino acid 169 to amino acid 178.

Other embodiments provide the gene corresponding to the cDNA sequence of SEQ ID NO:1.

In other embodiments, the present invention provides a composition comprising a protein, wherein said protein comprises an amino acid sequence selected from the group consisting of:

(a) the amino acid sequence of SEQ ID NO:2;

(b) the amino acid sequence of SEQ ID NO:2 from amino acid 169 to amino acid 178;

(c) fragments of the amino acid sequence of SEQ ID NO:2 comprising the amino acid sequence from amino acid 84 to amino acid 93 of SEQ ID NO:2; and

(d) the amino acid sequence encoded by the cDNA insert of clone en539_8 deposited under accession number ATCC 98408;

the protein being substantially free from other mammalian proteins. Preferably such protein comprises the amino acid sequence of SEQ ID NO:2 or the amino acid sequence of SEQ ID NO:2 from amino acid 169 to amino acid 178.

In one embodiment, the present invention provides a composition comprising an isolated polynucleotide selected from the group consisting of:

(a) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:3;

(b) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:3 from nucleotide 91 to nucleotide 966;

(c) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:3 from nucleotide 1 to nucleotide 337;

(d) a polynucleotide comprising the nucleotide sequence of the full-length protein coding sequence of clone eq188_1 deposited under accession number ATCC 98408;

(e) a polynucleotide encoding the full-length protein encoded by the cDNA insert of clone eq188_1 deposited under accession number ATCC 98408;

(f) a polynucleotide comprising the nucleotide sequence of a mature protein coding sequence of clone eq188_1 deposited under accession number ATCC 98408;

(g) a polynucleotide encoding a mature protein encoded by the cDNA insert of clone eq188_1 deposited under accession number ATCC 98408;

(h) a polynucleotide encoding a protein comprising the amino acid sequence of SEQ ID NO:4;

(i) a polynucleotide encoding a protein comprising a fragment of the amino acid sequence of SEQ ID NO:4 having biological activity, the fragment comprising the amino acid sequence from amino acid 141 to amino acid 150 of SEQ ID NO:4;

(j) a polynucleotide which is an allelic variant of a polynucleotide of (a)-(g) above;

(k) a polynucleotide which encodes a species homologue of the protein of (h) or (i) above ; and

(l) a polynucleotide that hybridizes under stringent conditions to any one of the polynucleotides specified in (a)-(i).

Preferably, such polynucleotide comprises the nucleotide sequence of SEQ ID NO:3 from nucleotide 91 to nucleotide 966; the nucleotide sequence of SEQ ID NO:3 from nucleotide 1 to nucleotide 337; the nucleotide sequence of the full-length protein coding sequence of clone eq188_1 deposited under accession number ATCC 98408; or the nucleotide sequence of a mature protein coding sequence of clone eq188_1 deposited under accession number ATCC 98408. In other preferred embodiments, the polynucleotide encodes the full-length or a mature protein encoded by the cDNA insert of clone eq188_1 deposited under accession number ATCC 98408. In yet other preferred embodiments, the present invention provides a polynucleotide encoding a protein comprising the amino acid sequence of SEQ ID NO:4 from amino acid 1 to amino acid 83.

Other embodiments provide the gene corresponding to the cDNA sequence of SEQ ID NO:3.

In other embodiments, the present invention provides a composition comprising a protein, wherein said protein comprises an amino acid sequence selected from the group consisting of:

(a) the amino acid sequence of SEQ ID NO:4;

(b) the amino acid sequence of SEQ ID NO:4 from amino acid 1 to amino acid 83;

(c) fragments of the amino acid sequence of SEQ ID NO:4 comprising the amino acid sequence from amino acid 141 to amino acid 150 of SEQ ID NO:4; and

(d) the amino acid sequence encoded by the cDNA insert of clone eq188_1 deposited under accession number ATCC 98408;

the protein being substantially free from other mammalian proteins. Preferably such protein comprises the amino acid sequence of SEQ ID NO:4 or the amino acid sequence of SEQ ID NO:4 from amino acid 1 to amino acid 83.

In one embodiment, the present invention provides a composition comprising an isolated polynucleotide selected from the group consisting of:

(a) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:5;

(b) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:5 from nucleotide 51 to nucleotide 1358;

(c) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:5 from nucleotide 99 to nucleotide 1358;

(d) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:5 from nucleotide 249 to nucleotide 566;

(e) a polynucleotide comprising the nucleotide sequence of the full-length protein coding sequence of clone er80_1 deposited under accession number ATCC 98408;

(f) a polynucleotide encoding the full-length protein encoded by the cDNA insert of clone er80_1 deposited under accession number ATCC 98408;

(g) a polynucleotide comprising the nucleotide sequence of a mature protein coding sequence of clone er80_1 deposited under accession number ATCC 98408;

(h) a polynucleotide encoding a mature protein encoded by the cDNA insert of clone er80_1 deposited under accession number ATCC 98408;

(i) a polynucleotide encoding a protein comprising the amino acid sequence of SEQ ID NO:6;

(j) a polynucleotide encoding a protein comprising a fragment of the amino acid sequence of SEQ ID NO:6 having biological activity, the fragment comprising the amino acid sequence from amino acid 213 to amino acid 222 of SEQ ID NO:6;

(k) a polynucleotide which is an allelic variant of a polynucleotide of (a)-(h) above;

(l) a polynucleotide which encodes a species homologue of the protein of (i) or (j) above ; and

(m) a polynucleotide that hybridizes under stringent conditions to any one of the polynucleotides specified in (a)-(j).

Preferably, such polynucleotide comprises the nucleotide sequence of SEQ ID NO:5 from nucleotide 51 to nucleotide 1358; the nucleotide sequence of SEQ ID NO:5 from nucleotide 99 to nucleotide 1358; the nucleotide sequence of SEQ ID NO:5 from nucleotide 249 to nucleotide 566; the nucleotide sequence of the full-length protein coding sequence of clone er80_1 deposited under accession number ATCC 98408; or the nucleotide sequence of a mature protein coding sequence of clone er80_1 deposited under accession number ATCC 98408. In other preferred embodiments, the polynucleotide encodes the full-length or a mature protein encoded by the cDNA insert of clone er80_1 deposited under accession number ATCC 98408. In yet other preferred embodiments, the present invention provides a polynucleotide encoding a protein comprising the amino acid sequence of SEQ ID NO:6 from amino acid 1 to amino acid 172.

Other embodiments provide the gene corresponding to the cDNA sequence of SEQ ID NO:5.

In other embodiments, the present invention provides a composition comprising a protein, wherein said protein comprises an amino acid sequence selected from the group consisting of:

(a) the amino acid sequence of SEQ ID NO:6;

(b) the amino acid sequence of SEQ ID NO:6 from amino acid 1 to amino acid 172;

(c) fragments of the amino acid sequence of SEQ ID NO:6 comprising the amino acid sequence from amino acid 213 to amino acid 222 of SEQ ID NO:6; and

(d) the amino acid sequence encoded by the cDNA insert of clone er80_1 deposited under accession number ATCC 98408;

the protein being substantially free from other mammalian proteins. Preferably such protein comprises the amino acid sequence of SEQ ID NO:6 or the amino acid sequence of SEQ ID NO:6 from amino acid 1 to amino acid 172.

In one embodiment, the present invention provides a composition comprising an isolated polynucleotide selected from the group consisting of:

(a) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:7;

(b) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:7 from nucleotide 571 to nucleotide 3306;

(c) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:7 from nucleotide 726 to nucleotide 1320;

(d) a polynucleotide comprising the nucleotide sequence of the full-length protein coding sequence of clone er418_5 deposited under accession number ATCC 98408;

(e) a polynucleotide encoding the full-length protein encoded by the cDNA insert of clone er418_5 deposited under accession number ATCC 98408;

(f) a polynucleotide comprising the nucleotide sequence of a mature protein coding sequence of clone er418_5 deposited under accession number ATCC 98408;

(g) a polynucleotide encoding a mature protein encoded by the cDNA insert of clone er418_5 deposited under accession number ATCC 98408;

(h) a polynucleotide encoding a protein comprising the amino acid sequence of SEQ ID NO:8;

(i) a polynucleotide encoding a protein comprising a fragment of the amino acid sequence of SEQ ID NO:8 having biological activity, the fragment comprising the amino acid sequence from amino acid 450 to amino acid 459 of SEQ ID NO:8;

(j) a polynucleotide which is an allelic variant of a polynucleotide of (a)-(g) above;

(k) a polynucleotide which encodes a species homologue of the protein of (h) or (i) above ; and

(l) a polynucleotide that hybridizes under stringent conditions to any one of the polynucleotides specified in (a)-(i).

Preferably, such polynucleotide comprises the nucleotide sequence of SEQ ID NO:7 from nucleotide 571 to nucleotide 3306; the nucleotide sequence of SEQ ID NO:7 from nucleotide 726 to nucleotide 1320; the nucleotide sequence of the full-length protein coding sequence of clone er418_5 deposited under accession number ATCC 98408; or the nucleotide sequence of a mature protein coding sequence of clone er418_5 deposited under accession number ATCC 98408. In other preferred embodiments, the polynucleotide encodes the full-length or a mature protein encoded by the cDNA insert of clone er418_5 deposited under accession number ATCC 98408. In yet other preferred embodiments, the present invention provides a polynucleotide encoding a protein comprising the amino acid sequence of SEQ ID NO:8 from amino acid 71 to amino acid 250.

Other embodiments provide the gene corresponding to the cDNA sequence of SEQ ID NO:7.

In other embodiments, the present invention provides a composition comprising a protein, wherein said protein comprises an amino acid sequence selected from the group consisting of:

(a) the amino acid sequence of SEQ ID NO:8;

(b) the amino acid sequence of SEQ ID NO:8 from amino acid 71 to amino acid 250;

(c) fragments of the amino acid sequence of SEQ ID NO:8 comprising the amino acid sequence from amino acid 450 to amino acid 459 of SEQ ID NO:8; and

(d) the amino acid sequence encoded by the cDNA insert of clone er418_5 deposited under accession number ATCC 98408;

the protein being substantially free from other mammalian proteins. Preferably such protein comprises the amino acid sequence of SEQ ID NO:8 or the amino acid sequence of SEQ ID NO:8 from amino acid 71 to amino acid 250.

In one embodiment, the present invention provides a composition comprising an isolated polynucleotide selected from the group consisting of:

(a) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:9;

(b) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:9 from nucleotide 503 to nucleotide 2770;

(c) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:9 from nucleotide 572 to nucleotide 2770;

(d) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:9 from nucleotide 490 to nucleotide 772;

(e) a polynucleotide comprising the nucleotide sequence of the full-length protein coding sequence of clone fa252_8 deposited under accession number ATCC 98408;

(f) a polynucleotide encoding the full-length protein encoded by the cDNA insert of clone fa252_8 deposited under accession number ATCC 98408;

(g) a polynucleotide comprising the nucleotide sequence of a mature protein coding sequence of clone fa252_8 deposited under accession number ATCC 98408;

(h) a polynucleotide encoding a mature protein encoded by the cDNA insert of clone fa252_8 deposited under accession number ATCC 98408;

(i) a polynucleotide encoding a protein comprising the amino acid sequence of SEQ ID NO:10;

(j) a polynucleotide encoding a protein comprising a fragment of the amino acid sequence of SEQ ID NO:10 having biological activity, the fragment comprising the amino acid sequence from amino acid 373 to amino acid 382 of SEQ ID NO:10;

(k) a polynucleotide which is an allelic variant of a polynucleotide of (a)-(h) above;

(l) a polynucleotide which encodes a species homologue of the protein of (i) or (j) above ; and

(m) a polynucleotide that hybridizes under stringent conditions to any one of the polynucleotides specified in (a)-(j).

Preferably, such polynucleotide comprises the nucleotide sequence of SEQ ID NO:9 from nucleotide 503 to nucleotide 2770; the nucleotide sequence of SEQ ID NO:9 from nucleotide 572 to nucleotide 2770; the nucleotide sequence of SEQ ID NO:9 from nucleotide 490 to nucleotide 772; the nucleotide sequence of the full-length protein coding

sequence of clone fa252_8 deposited under accession number ATCC 98408; or the nucleotide sequence of a mature protein coding sequence of clone fa252_8 deposited under accession number ATCC 98408. In other preferred embodiments, the polynucleotide encodes the full-length or a mature protein encoded by the cDNA insert of clone fa252_8 deposited under accession number ATCC 98408. In yet other preferred embodiments, the present invention provides a polynucleotide encoding a protein comprising the amino acid sequence of SEQ ID NO:10 from amino acid 1 to amino acid 90.

Other embodiments provide the gene corresponding to the cDNA sequence of SEQ ID NO:9.

In other embodiments, the present invention provides a composition comprising a protein, wherein said protein comprises an amino acid sequence selected from the group consisting of:

(a) the amino acid sequence of SEQ ID NO:10;

(b) the amino acid sequence of SEQ ID NO:10 from amino acid 1 to amino acid 90;

(c) fragments of the amino acid sequence of SEQ ID NO:10 comprising the amino acid sequence from amino acid 373 to amino acid 382 of SEQ ID NO:10; and

(d) the amino acid sequence encoded by the cDNA insert of clone fa252_8 deposited under accession number ATCC 98408;

the protein being substantially free from other mammalian proteins. Preferably such protein comprises the amino acid sequence of SEQ ID NO:10 or the amino acid sequence of SEQ ID NO:10 from amino acid 1 to amino acid 90.

In one embodiment, the present invention provides a composition comprising an isolated polynucleotide selected from the group consisting of:

(a) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:11;

(b) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:11 from nucleotide 104 to nucleotide 565;

(c) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:11 from nucleotide 1 to nucleotide 501;

(d) a polynucleotide comprising the nucleotide sequence of the full-length protein coding sequence of clone fg912_1 deposited under accession number ATCC 98408;

(e) a polynucleotide encoding the full-length protein encoded by the cDNA insert of clone fg912_1 deposited under accession number ATCC 98408;

(f) a polynucleotide comprising the nucleotide sequence of a mature protein coding sequence of clone fg912_1 deposited under accession number ATCC 98408;

(g) a polynucleotide encoding a mature protein encoded by the cDNA insert of clone fg912_1 deposited under accession number ATCC 98408;

(h) a polynucleotide encoding a protein comprising the amino acid sequence of SEQ ID NO:12;

(i) a polynucleotide encoding a protein comprising a fragment of the amino acid sequence of SEQ ID NO:12 having biological activity, the fragment comprising the amino acid sequence from amino acid 72 to amino acid 81 of SEQ ID NO:12;

(j) a polynucleotide which is an allelic variant of a polynucleotide of (a)-(g) above;

(k) a polynucleotide which encodes a species homologue of the protein of (h) or (i) above ; and

(l) a polynucleotide that hybridizes under stringent conditions to any one of the polynucleotides specified in (a)-(i).

Preferably, such polynucleotide comprises the nucleotide sequence of SEQ ID NO:11 from nucleotide 104 to nucleotide 565; the nucleotide sequence of SEQ ID NO:11 from nucleotide 1 to nucleotide 501; the nucleotide sequence of the full-length protein coding sequence of clone fg912_1 deposited under accession number ATCC 98408; or the nucleotide sequence of a mature protein coding sequence of clone fg912_1 deposited under accession number ATCC 98408. In other preferred embodiments, the polynucleotide encodes the full-length or a mature protein encoded by the cDNA insert of clone fg912_1 deposited under accession number ATCC 98408. In yet other preferred embodiments, the present invention provides a polynucleotide encoding a protein comprising the amino acid sequence of SEQ ID NO:12 from amino acid 1 to amino acid 132.

Other embodiments provide the gene corresponding to the cDNA sequence of SEQ ID NO:11.

In other embodiments, the present invention provides a composition comprising a protein, wherein said protein comprises an amino acid sequence selected from the group consisting of:

(a) the amino acid sequence of SEQ ID NO:12;

(b) the amino acid sequence of SEQ ID NO:12 from amino acid 1 to amino acid 132;

(c) fragments of the amino acid sequence of SEQ ID NO:12 comprising the amino acid sequence from amino acid 72 to amino acid 81 of SEQ ID NO:12; and

(d) the amino acid sequence encoded by the cDNA insert of clone fg912_1 deposited under accession number ATCC 98408;

the protein being substantially free from other mammalian proteins. Preferably such protein comprises the amino acid sequence of SEQ ID NO:12 or the amino acid sequence of SEQ ID NO:12 from amino acid 1 to amino acid 132.

In one embodiment, the present invention provides a composition comprising an isolated polynucleotide selected from the group consisting of:

(a) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:13;

(b) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:13 from nucleotide 77 to nucleotide 1093;

(c) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:13 from nucleotide 167 to nucleotide 1093;

(d) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:13 from nucleotide 1 to nucleotide 718;

(e) a polynucleotide comprising the nucleotide sequence of the full-length protein coding sequence of clone fg949_3 deposited under accession number ATCC 98408;

(f) a polynucleotide encoding the full-length protein encoded by the cDNA insert of clone fg949_3 deposited under accession number ATCC 98408;

(g) a polynucleotide comprising the nucleotide sequence of a mature protein coding sequence of clone fg949_3 deposited under accession number ATCC 98408;

(h) a polynucleotide encoding a mature protein encoded by the cDNA insert of clone fg949_3 deposited under accession number ATCC 98408;

(i) a polynucleotide encoding a protein comprising the amino acid sequence of SEQ ID NO:14;

(j) a polynucleotide encoding a protein comprising a fragment of the amino acid sequence of SEQ ID NO:14 having biological activity, the fragment comprising the amino acid sequence from amino acid 164 to amino acid 173 of SEQ ID NO:14;

(k) a polynucleotide which is an allelic variant of a polynucleotide of (a)-(h) above;

(l) a polynucleotide which encodes a species homologue of the protein of (i) or (j) above ; and

(m) a polynucleotide that hybridizes under stringent conditions to any one of the polynucleotides specified in (a)-(j).

Preferably, such polynucleotide comprises the nucleotide sequence of SEQ ID NO:13 from nucleotide 77 to nucleotide 1093; the nucleotide sequence of SEQ ID NO:13 from nucleotide 167 to nucleotide 1093; the nucleotide sequence of SEQ ID NO:13 from nucleotide 1 to nucleotide 718; the nucleotide sequence of the full-length protein coding sequence of clone fg949_3 deposited under accession number ATCC 98408; or the nucleotide sequence of a mature protein coding sequence of clone fg949_3 deposited under accession number ATCC 98408. In other preferred embodiments, the polynucleotide encodes the full-length or a mature protein encoded by the cDNA insert of clone fg949_3 deposited under accession number ATCC 98408. In yet other preferred embodiments, the present invention provides a polynucleotide encoding a protein comprising the amino acid sequence of SEQ ID NO:14 from amino acid 1 to amino acid 214.

Other embodiments provide the gene corresponding to the cDNA sequence of SEQ ID NO:13.

In other embodiments, the present invention provides a composition comprising a protein, wherein said protein comprises an amino acid sequence selected from the group consisting of:

(a) the amino acid sequence of SEQ ID NO:14;

(b) the amino acid sequence of SEQ ID NO:14 from amino acid 1 to amino acid 214;

(c) fragments of the amino acid sequence of SEQ ID NO:14 comprising the amino acid sequence from amino acid 164 to amino acid 173 of SEQ ID NO:14; and

(d) the amino acid sequence encoded by the cDNA insert of clone fg949_3 deposited under accession number ATCC 98408;

the protein being substantially free from other mammalian proteins. Preferably such protein comprises the amino acid sequence of SEQ ID NO:14 or the amino acid sequence of SEQ ID NO:14 from amino acid 1 to amino acid 214.

In one embodiment, the present invention provides a composition comprising an isolated polynucleotide selected from the group consisting of:

(a) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:15;

(b) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:15 from nucleotide 19 to nucleotide 1023;

(c) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:15 from nucleotide 247 to nucleotide 711;

(d) a polynucleotide comprising the nucleotide sequence of the full-length protein coding sequence of clone fk354_4 deposited under accession number ATCC 98408;

(e) a polynucleotide encoding the full-length protein encoded by the cDNA insert of clone fk354_4 deposited under accession number ATCC 98408;

(f) a polynucleotide comprising the nucleotide sequence of a mature protein coding sequence of clone fk354_4 deposited under accession number ATCC 98408;

(g) a polynucleotide encoding a mature protein encoded by the cDNA insert of clone fk354_4 deposited under accession number ATCC 98408;

(h) a polynucleotide encoding a protein comprising the amino acid sequence of SEQ ID NO:16;

(i) a polynucleotide encoding a protein comprising a fragment of the amino acid sequence of SEQ ID NO:16 having biological activity, the fragment comprising the amino acid sequence from amino acid 162 to amino acid 171 of SEQ ID NO:16;

(j) a polynucleotide which is an allelic variant of a polynucleotide of (a)-(g) above;

(k) a polynucleotide which encodes a species homologue of the protein of (h) or (i) above ; and

(l) a polynucleotide that hybridizes under stringent conditions to any one of the polynucleotides specified in (a)-(i).

Preferably, such polynucleotide comprises the nucleotide sequence of SEQ ID NO:15 from nucleotide 19 to nucleotide 1023; the nucleotide sequence of SEQ ID NO:15 from nucleotide 247 to nucleotide 711; the nucleotide sequence of the full-length protein coding sequence of clone fk354_4 deposited under accession number ATCC 98408; or the nucleotide sequence of a mature protein coding sequence of clone fk354_4 deposited under accession number ATCC 98408. In other preferred embodiments, the polynucleotide encodes the full-length or a mature protein encoded by the cDNA insert of clone fk354_4 deposited under accession number ATCC 98408. In yet other preferred embodiments, the present invention provides a polynucleotide encoding a protein comprising the amino acid sequence of SEQ ID NO:16 from amino acid 147 to amino acid 231.

Other embodiments provide the gene corresponding to the cDNA sequence of SEQ ID NO:15.

In other embodiments, the present invention provides a composition comprising a protein, wherein said protein comprises an amino acid sequence selected from the group consisting of:

(a) the amino acid sequence of SEQ ID NO:16;

(b) the amino acid sequence of SEQ ID NO:16 from amino acid 147 to amino acid 231;

(c) fragments of the amino acid sequence of SEQ ID NO:16 comprising the amino acid sequence from amino acid 162 to amino acid 171 of SEQ ID NO:16; and

(d) the amino acid sequence encoded by the cDNA insert of clone fk354_4 deposited under accession number ATCC 98408;

the protein being substantially free from other mammalian proteins. Preferably such protein comprises the amino acid sequence of SEQ ID NO:16 or the amino acid sequence of SEQ ID NO:16 from amino acid 147 to amino acid 231.

In one embodiment, the present invention provides a composition comprising an isolated polynucleotide selected from the group consisting of:

(a) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:17;

(b) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:17 from nucleotide 11 to nucleotide 970;

(c) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:17 from nucleotide 1 to nucleotide 575;

(d) a polynucleotide comprising the nucleotide sequence of the full-length protein coding sequence of clone fm150_1 deposited under accession number ATCC 98408;

(e) a polynucleotide encoding the full-length protein encoded by the cDNA insert of clone fm150_1 deposited under accession number ATCC 98408;

(f) a polynucleotide comprising the nucleotide sequence of a mature protein coding sequence of clone fm150_1 deposited under accession number ATCC 98408;

(g) a polynucleotide encoding a mature protein encoded by the cDNA insert of clone fm150_1 deposited under accession number ATCC 98408;

(h) a polynucleotide encoding a protein comprising the amino acid sequence of SEQ ID NO:18;

(i) a polynucleotide encoding a protein comprising a fragment of the amino acid sequence of SEQ ID NO:18 having biological activity, the fragment comprising the amino acid sequence from amino acid 155 to amino acid 164 of SEQ ID NO:18;

(j) a polynucleotide which is an allelic variant of a polynucleotide of (a)-(g) above;

(k) a polynucleotide which encodes a species homologue of the protein of (h) or (i) above ; and

(l) a polynucleotide that hybridizes under stringent conditions to any one of the polynucleotides specified in (a)-(i).

Preferably, such polynucleotide comprises the nucleotide sequence of SEQ ID NO:17 from nucleotide 11 to nucleotide 970; the nucleotide sequence of SEQ ID NO:17 from nucleotide 1 to nucleotide 575; the nucleotide sequence of the full-length protein coding sequence of clone fm150_1 deposited under accession number ATCC 98408; or the nucleotide sequence of a mature protein coding sequence of clone fm150_1 deposited under accession number ATCC 98408. In other preferred embodiments, the

polynucleotide encodes the full-length or a mature protein encoded by the cDNA insert of clone fm150_1 deposited under accession number ATCC 98408. In yet other preferred embodiments, the present invention provides a polynucleotide encoding a protein comprising the amino acid sequence of SEQ ID NO:18 from amino acid 1 to amino acid 188.

Other embodiments provide the gene corresponding to the cDNA sequence of SEQ ID NO:17.

In other embodiments, the present invention provides a composition comprising a protein, wherein said protein comprises an amino acid sequence selected from the group consisting of:

(a) the amino acid sequence of SEQ ID NO:18;

(b) the amino acid sequence of SEQ ID NO:18 from amino acid 1 to amino acid 188;

(c) fragments of the amino acid sequence of SEQ ID NO:18 comprising the amino acid sequence from amino acid 155 to amino acid 164 of SEQ ID NO:18; and

(d) the amino acid sequence encoded by the cDNA insert of clone fm150_1 deposited under accession number ATCC 98408;

the protein being substantially free from other mammalian proteins. Preferably such protein comprises the amino acid sequence of SEQ ID NO:18 or the amino acid sequence of SEQ ID NO:18 from amino acid 1 to amino acid 188.

In one embodiment, the present invention provides a composition comprising an isolated polynucleotide selected from the group consisting of:

(a) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:19;

(b) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:19 from nucleotide 223 to nucleotide 882;

(c) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:19 from nucleotide 46 to nucleotide 351;

(d) a polynucleotide comprising the nucleotide sequence of the full-length protein coding sequence of clone gu534_1 deposited under accession number ATCC 98408;

(e) a polynucleotide encoding the full-length protein encoded by the cDNA insert of clone gu534_1 deposited under accession number ATCC 98408;

(f) a polynucleotide comprising the nucleotide sequence of a mature protein coding sequence of clone gu534_1 deposited under accession number ATCC 98408;

(g) a polynucleotide encoding a mature protein encoded by the cDNA insert of clone gu534_1 deposited under accession number ATCC 98408;

(h) a polynucleotide encoding a protein comprising the amino acid sequence of SEQ ID NO:20;

(i) a polynucleotide encoding a protein comprising a fragment of the amino acid sequence of SEQ ID NO:20 having biological activity, the fragment comprising the amino acid sequence from amino acid 105 to amino acid 114 of SEQ ID NO:20;

(j) a polynucleotide which is an allelic variant of a polynucleotide of (a)-(g) above;

(k) a polynucleotide which encodes a species homologue of the protein of (h) or (i) above ; and

(l) a polynucleotide that hybridizes under stringent conditions to any one of the polynucleotides specified in (a)-(i).

Preferably, such polynucleotide comprises the nucleotide sequence of SEQ ID NO:19 from nucleotide 223 to nucleotide 882; the nucleotide sequence of SEQ ID NO:19 from nucleotide 46 to nucleotide 351; the nucleotide sequence of the full-length protein coding sequence of clone gu534_1 deposited under accession number ATCC 98408; or the nucleotide sequence of a mature protein coding sequence of clone gu534_1 deposited under accession number ATCC 98408. In other preferred embodiments, the polynucleotide encodes the full-length or a mature protein encoded by the cDNA insert of clone gu534_1 deposited under accession number ATCC 98408. In yet other preferred embodiments, the present invention provides a polynucleotide encoding a protein comprising the amino acid sequence of SEQ ID NO:20 from amino acid 1 to amino acid 43.

Other embodiments provide the gene corresponding to the cDNA sequence of SEQ ID NO:19.

In other embodiments, the present invention provides a composition comprising a protein, wherein said protein comprises an amino acid sequence selected from the group consisting of:

(a) the amino acid sequence of SEQ ID NO:20;

(b) the amino acid sequence of SEQ ID NO:20 from amino acid 1 to amino acid 43;

(c) fragments of the amino acid sequence of SEQ ID NO:20 comprising the amino acid sequence from amino acid 105 to amino acid 114 of SEQ ID NO:20; and

(d) the amino acid sequence encoded by the cDNA insert of clone gu534_1 deposited under accession number ATCC 98408;

the protein being substantially free from other mammalian proteins. Preferably such protein comprises the amino acid sequence of SEQ ID NO:20 or the amino acid sequence of SEQ ID NO:20 from amino acid 1 to amino acid 43.

In certain preferred embodiments, the polynucleotide is operably linked to an expression control sequence. The invention also provides a host cell, including bacterial, yeast, insect and mammalian cells, transformed with such polynucleotide compositions. Also provided by the present invention are organisms that have enhanced, reduced, or modified expression of the gene(s) corresponding to the polynucleotide sequences disclosed herein.

Processes are also provided for producing a protein, which comprise:

(a) growing a culture of the host cell transformed with such polynucleotide compositions in a suitable culture medium; and

(b) purifying the protein from the culture.

The protein produced according to such methods is also provided by the present invention. Preferred embodiments include those in which the protein produced by such process is a mature form of the protein.

Protein compositions of the present invention may further comprise a pharmaceutically acceptable carrier. Compositions comprising an antibody which specifically reacts with such protein are also provided by the present invention.

Methods are also provided for preventing, treating or ameliorating a medical condition which comprises administering to a mammalian subject a therapeutically effective amount of a composition comprising a protein of the present invention and a pharmaceutically acceptable carrier.

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1A and 1B are schematic representations of the pED6 and pNOTs vectors, respectively, used for deposit of clones disclosed herein.

DETAILED DESCRIPTION

ISOLATED PROTEINS AND POLYNUCLEOTIDES

Nucleotide and amino acid sequences, as presently determined, are reported below for each clone and protein disclosed in the present application. The nucleotide sequence of each clone can readily be determined by sequencing of the deposited clone in accordance with known methods. The predicted amino acid sequence (both full-length and mature forms) can then be determined from such nucleotide sequence. The amino acid sequence of the protein encoded by a particular clone can also be determined by expression of the clone in a suitable host cell, collecting the protein and determining its sequence. For each disclosed protein applicants have identified what they have determined to be the reading frame best identifiable with sequence information available at the time of filing.

As used herein a "secreted" protein is one which, when expressed in a suitable host cell, is transported across or through a membrane, including transport as a result of signal sequences in its amino acid sequence. "Secreted" proteins include without limitation proteins secreted wholly (e.g., soluble proteins) or partially (e.g. , receptors) from the cell in which they are expressed. "Secreted" proteins also include without limitation proteins which are transported across the membrane of the endoplasmic reticulum.

Clone "en539_8"

A polynucleotide of the present invention has been identified as clone "en539_8". en539_8 was isolated from a human fetal brain cDNA library using methods which are selective for cDNAs encoding secreted proteins (see U.S. Pat. No. 5,536,637), or was identified as encoding a secreted or transmembrane protein on the basis of computer analysis of the amino acid sequence of the encoded protein. en539_8 is a full-length clone, including the entire coding sequence of a secreted protein (also referred to herein as "en539_8 protein").

The nucleotide sequence of en539_8 as presently determined is reported in SEQ ID NO:1. What applicants presently believe to be the proper reading frame and the predicted amino acid sequence of the en539_8 protein corresponding to the foregoing nucleotide sequence is reported in SEQ ID NO:2. Amino acids 151 to 163 are a predicted leader/signal sequence, with the predicted mature amino acid sequence beginning at amino acid 164, or are a transmembrane domain.

The EcoRI/NotI restriction fragment obtainable from the deposit containing clone en539_8 should be approximately 2700 bp.

The nucleotide sequence disclosed herein for en539_8 was searched against the GenBank and GeneSeq nucleotide sequence databases using BLASTN/BLASTX and FASTA search protocols. en539_8 demonstrated at least some similarity with sequences identified as AC000353 (Homo sapiens chromosome 11 clone 18h3 from q13; HTGS phase 1, 14 unordered pieces), R80149 (yi95d12.s1 Homo sapiens cDNA clone), T54084 (ya92a05.s1 Homo sapiens cDNA clone 69104 3' contains L1 repetitive element), U07562 (Human ABL gene, intron 1b, partial sequence), and Z68886 (Human DNA sequence from cosmid L21F12, Huntington's Disease Region, chromosome 4p16.3). Based upon sequence similarity, en539_8 proteins and each similar protein or peptide may share at least some activity. The nucleotide sequence of en539_8 indicates that it may contain an Alu repetitive element.

Clone "eq188_1"

A polynucleotide of the present invention has been identified as clone "eq188_1". eq188_1 was isolated from a human adult testes cDNA library using methods which are selective for cDNAs encoding secreted proteins (see U.S. Pat. No. 5,536,637), or was identified as encoding a secreted or transmembrane protein on the basis of computer analysis of the amino acid sequence of the encoded protein. eq188_1 is a full-length clone, including the entire coding sequence of a secreted protein (also referred to herein as "eq188_1 protein").

The nucleotide sequence of eq188_1 as presently determined is reported in SEQ ID NO:3. What applicants presently believe to be the proper reading frame and the predicted amino acid sequence of the eq188_1 protein corresponding to the foregoing nucleotide sequence is reported in SEQ ID NO:4.

The EcoRI/NotI restriction fragment obtainable from the deposit containing clone eq188_1 should be approximately 1650 bp.

The nucleotide sequence disclosed herein for eq188_1 was searched against the GenBank and GeneSeq nucleotide sequence databases using BLASTN/BLASTX and FASTA search protocols. eq188_1 demonstrated at least some similarity with sequences identified as W31185 (zb87h03.r1 Soares senescent fibroblasts NbHSF Homo sapiens cDNA clone 310613 5). The predicted amino acid sequence disclosed herein for eq188_1 was searched against the GenPept and GeneSeq amino acid sequence databases using the

BLASTX search protocol. The predicted eq188_1 protein demonstrated at least some similarity to sequences identified as X85105 (spindle pole body protein [Schizosaccharomyces pombe]). Based upon sequence similarity, eq188_1 proteins and each similar protein or peptide may share at least some activity. The TopPredII computer program predicts a potential transmembrane domain within the eq188_1 protein sequence centered around amino acid 55 of SEQ ID NO:4.

Clone "er80_1"

A polynucleotide of the present invention has been identified as clone "er80_1". er80_1 was isolated from a human fetal brain cDNA library using methods which are selective for cDNAs encoding secreted proteins (see U.S. Pat. No. 5,536,637), or was identified as encoding a secreted or transmembrane protein on the basis of computer analysis of the amino acid sequence of the encoded protein. er80_1 is a full-length clone, including the entire coding sequence of a secreted protein (also referred to herein as "er80_1 protein").

The nucleotide sequence of er80_1 as presently determined is reported in SEQ ID NO:5. What applicants presently believe to be the proper reading frame and the predicted amino acid sequence of the er80_1 protein corresponding to the foregoing nucleotide sequence is reported in SEQ ID NO:6. Amino acids 4 to 16 are a predicted leader/signal sequence, with the predicted mature amino acid sequence beginning at amino acid 17.

The EcoRI/NotI restriction fragment obtainable from the deposit containing clone er80_1 should be approximately 3000 bp.

The nucleotide sequence disclosed herein for er80_1 was searched against the GenBank and GeneSeq nucleotide sequence databases using BLASTN/BLASTX and FASTA search protocols. er80_1 demonstrated at least some similarity with sequences identified as AA027861 (zk05a02.r1 Soares pregnant uterus NbHPU Homo sapiens cDNA clone 469610 5' similar to PIR S33293 S33293 testican - human), N47945 (yy84c11.s1 Homo sapiens cDNA clone 280244 3'), N77555 (yz89e09.r1 Homo sapiens cDNA clone 290248 5'), X73608 (H.sapiens mRNA for testican), and X92864 (M.musculus mRNA for testican). The predicted amino acid sequence disclosed herein for er80_1 was searched against the GenPept and GeneSeq amino acid sequence databases using the BLASTX search protocol. The predicted er80_1 protein demonstrated at least some similarity to sequences identified as X73608 (testican [Homo sapiens]). The predicted er80_1 protein contains the thyroglobulin type-1 repeat signature. Thyroglobulin (Tg) is a large glycoprotein specific

to the thyroid gland and is the precursor of the iodinated thyroid hormones thyroxine (T4) and triiodothyronine (T3). The N-terminal section of Tg contains ten repeats of a domain of about 65 amino acids which is known as the Tg type-1 repeat. This motif is also found in various cell surface and secreted proteins as a single copy, and it is found as a single copy in er80_1 protein. For example, in the HLA class II associated invariant chain, the Tg type-1 repeat is encoded by an exon which is alternatively spliced and is only present in a longer form of the protein, indicating that this motif has functional significance. Based upon sequence similarity, er80_1 proteins and each similar protein or peptide may share at least some activity.

Clone "er418_5"

A polynucleotide of the present invention has been identified as clone "er418_5". er418_5 was isolated from a human fetal brain cDNA library using methods which are selective for cDNAs encoding secreted proteins (see U.S. Pat. No. 5,536,637), or was identified as encoding a secreted or transmembrane protein on the basis of computer analysis of the amino acid sequence of the encoded protein. er418_5 is a full-length clone, including the entire coding sequence of a secreted protein (also referred to herein as "er418_5 protein").

The nucleotide sequence of er418_5 as presently determined is reported in SEQ ID NO:7. What applicants presently believe to be the proper reading frame and the predicted amino acid sequence of the er418_5 protein corresponding to the foregoing nucleotide sequence is reported in SEQ ID NO:8.

The EcoRI/NotI restriction fragment obtainable from the deposit containing clone er418_5 should be approximately 3800 bp.

The nucleotide sequence disclosed herein for er418_5 was searched against the GenBank and GeneSeq nucleotide sequence databases using BLASTN/BLASTX and FASTA search protocols. er418_5 demonstrated at least some similarity with sequences identified as AA024596 (ze78a11.s1 Soares fetal heart NbHH19W Homo sapiens cDNA clone 365084 3'), AA181258 (zp58d01.s1 Stratagene endothelial cell 937223 Homo sapiens cDNA clone 624385 3'), Q39674 (Expressed Sequence Tag human gene marker EST00046), W28438 (47g10 Human retina cDNA randomly primed sublibrary Homo sapiens cDNA), and Z36842 (H.sapiens (xs85) mRNA, 209bp). The predicted amino acid sequence disclosed herein for er418_5 was searched against the GenPept and GeneSeq amino acid sequence databases using the BLASTX search protocol. The predicted er418_5 protein

demonstrated at least some similarity to sequences identified as M80902 (AHNAK nucleoprotein [Homo sapiens]). Based upon sequence similarity, er418_5 proteins and each similar protein or peptide may share at least some activity. The TopPredII computer program predicts a potential transmembrane domain within the er418_5 protein sequence centered around amino acid 760 of SEQ ID NO:8.

Clone "fa252_8"

A polynucleotide of the present invention has been identified as clone "fa252_8". fa252_8 was isolated from a human fetal brain cDNA library using methods which are selective for cDNAs encoding secreted proteins (see U.S. Pat. No. 5,536,637), or was identified as encoding a secreted or transmembrane protein on the basis of computer analysis of the amino acid sequence of the encoded protein. fa252_8 is a full-length clone, including the entire coding sequence of a secreted protein (also referred to herein as "fa252_8 protein").

The nucleotide sequence of fa252_8 as presently determined is reported in SEQ ID NO:9. What applicants presently believe to be the proper reading frame and the predicted amino acid sequence of the fa252_8 protein corresponding to the foregoing nucleotide sequence is reported in SEQ ID NO:10. Amino acids 11 to 23 are a predicted leader/signal sequence, with the predicted mature amino acid sequence beginning at amino acid 24, or are a transmembrane domain.

The EcoRI/NotI restriction fragment obtainable from the deposit containing clone fa252_8 should be approximately 4300 bp.

The nucleotide sequence disclosed herein for fa252_8 was searched against the GenBank and GeneSeq nucleotide sequence databases using BLASTN/BLASTX and FASTA search protocols. fa252_8 demonstrated at least some similarity with sequences identified as AA001054 (ze47e04.s1 Soares retina N2b4HR Homo sapiens cDNA clone 362142 3'), AA029283 (zk10a03.s1 Soares pregnant uterus NbHPU Homo sapiens cDNA clone 470092 3'), AL008630 (Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone 282F2; HTGS phase 1), Z68287 (Human DNA sequence from cosmid N38E12, between markers D22S280 and D22S86 on chromosome 22q12), Z69042 (Human DNA sequence from cosmid E95B1, between markers D22S280 and D22S86 on chromosome 22q12), and Z73429 Human DNA sequence from cosmid cN32F9 on chromosome 22q11.2-qter Contains CpG island). The predicted amino acid sequence disclosed herein for fa252_8 was searched against the GenPept and GeneSeq amino acid sequence

databases using the BLASTX search protocol. The predicted fa252_8 protein demonstrated at least some similarity to sequences identified as D14157 (calcium channel BIII [Oryctolagus cuniculus]) and Z68006 (K09C8.4 [Caenorhabditis elegans]). Based upon sequence similarity, fa252_8 proteins and each similar protein or peptide may share at least some activity. The TopPredII computer program predicts an additional potential transmembrane domain within the fa252_8 protein sequence centered around amino acid 190 of SEQ ID NO:10.

Clone "fg912_1"

A polynucleotide of the present invention has been identified as clone "fg912_1". fg912_1 was isolated from a human adult brain cDNA library using methods which are selective for cDNAs encoding secreted proteins (see U.S. Pat. No. 5,536,637), or was identified as encoding a secreted or transmembrane protein on the basis of computer analysis of the amino acid sequence of the encoded protein. fg912_1 is a full-length clone, including the entire coding sequence of a secreted protein (also referred to herein as "fg912_1 protein").

The nucleotide sequence of fg912_1 as presently determined is reported in SEQ ID NO:11. What applicants presently believe to be the proper reading frame and the predicted amino acid sequence of the fg912_1 protein corresponding to the foregoing nucleotide sequence is reported in SEQ ID NO:12.

The EcoRI/NotI restriction fragment obtainable from the deposit containing clone fg912_1 should be approximately 1800 bp.

The nucleotide sequence disclosed herein for fg912_1 was searched against the GenBank and GeneSeq nucleotide sequence databases using BLASTN/BLASTX and FASTA search protocols. fg912_1 demonstrated at least some similarity with sequences identified as AA043948 (zk58c06.r1 Soares pregnant uterus NbHPU Homo sapiens cDNA clone 487018 5'), AA081739 (zn23c06.r1 Stratagene neuroepithelium NT2RAMI 937234 Homo sapiens cDNA clone 548266 5'), AA114831 (zk88e07.s1 Soares pregnant uterus NbHPU Homo sapiens cDNA clone 489924 3'), AA151779 (zo39e10.r1 Stratagene endothelial cell 937223 Homo sapiens cDNA clone 589290 5'), AA205696 (zq69h08.s1 Stratagene neuroepithelium (#937231) Homo sapiens cDNA clone 646911 3'), N34239 (yx79c05.r1 Homo sapiens cDNA clone 267944 5'), R59637 (yh02a07.r1 Homo sapiens cDNA clone 41898 5'), T24418 (Human gene signature HUMGS06451), T26513 (Human gene signature HUMGS08755), T35507 (EST86582 Homo sapiens cDNA 5' end similar to

None), and U90123 (Mus musculus HN1 (Hn1) mRNA, complete cds). The predicted amino acid sequence disclosed herein for fg912_1 was searched against the GenPept and GeneSeq amino acid sequence databases using the BLASTX search protocol. The predicted fg912_1 protein demonstrated at least some similarity to sequences identified as U90123 (HN1 [Mus musculus]). Based upon sequence similarity, fg912_1 proteins and each similar protein or peptide may share at least some activity.

Clone "fg949_3"

A polynucleotide of the present invention has been identified as clone "fg949_3". fg949_3 was isolated from a human adult brain cDNA library using methods which are selective for cDNAs encoding secreted proteins (see U.S. Pat. No. 5,536,637), or was identified as encoding a secreted or transmembrane protein on the basis of computer analysis of the amino acid sequence of the encoded protein. fg949_3 is a full-length clone, including the entire coding sequence of a secreted protein (also referred to herein as "fg949_3 protein").

The nucleotide sequence of fg949_3 as presently determined is reported in SEQ ID NO:13. What applicants presently believe to be the proper reading frame and the predicted amino acid sequence of the fg949_3 protein corresponding to the foregoing nucleotide sequence is reported in SEQ ID NO:14. Amino acids 18 to 30 are a predicted leader/signal sequence, with the predicted mature amino acid sequence beginning at amino acid 31, or are a transmembrane domain.

The EcoRI/NotI restriction fragment obtainable from the deposit containing clone fg949_3 should be approximately 2200 bp.

The nucleotide sequence disclosed herein for fg949_3 was searched against the GenBank and GeneSeq nucleotide sequence databases using BLASTN/BLASTX and FASTA search protocols. fg949_3 demonstrated at least some similarity with sequences identified as AA001371 (ze45a04.s1 Soares retina N2b4HR Homo sapiens cDNA clone 361902 3'), AA059397 (zf67f10.s1 Soares pineal gland N3HPG Homo sapiens cDNA clone 382027 3'), AA084199 (zn17e04.r1 Stratagene neuroepithelium NT2RAMI 937234 Homo sapiens cDNA clone 547710 5' similar to WP:T06D8.9 CE02330), H51759 (yp81f10.r1 Homo sapiens cDNA clone 193867 5'), H53493 (yq86e01.r1 Homo sapiens cDNA clone 202680 5'), T22173 (Human gene signature HUMGS03744), T31244 (EST29112 Homo sapiens cDNA 5' end similar to None), T82823 (yd38e02.r1 Homo sapiens cDNA clone 110522 5'), W02871 (za05e06.r1 Soares melanocyte 2NbHM Homo sapiens cDNA clone 291682 5' similar to

WP T06D8.9 CE02330), W19556 (zb31c04.r1 Soares parathyroid tumor NbHPA Homo sapiens cDNA clone 305190 5' similar to WP:T06D8.9 CE02330), and Z70223 (H.sapiens mRNA for 5'UTR for unknown protein (clone ICRFp507L0677)). The predicted amino acid sequence disclosed herein for fg949_3 was searched against the GenPept and GeneSeq amino acid sequence databases using the BLASTX search protocol. The predicted fg949_3 protein demonstrated at least some similarity to sequences identified as Z49130 (T06D8.9 [Caenorhabditis elegans]). Based upon sequence similarity, fg949_3 proteins and each similar protein or peptide may share at least some activity. The TopPredII computer program predicts an additional potential transmembrane domain within the fg949_3 protein sequence centered around amino acid 180 of SEQ ID NO:14.

Clone "fk354_4"

A polynucleotide of the present invention has been identified as clone "fk354_4". fk354_4 was isolated from a human adult kidney cDNA library using methods which are selective for cDNAs encoding secreted proteins (see U.S. Pat. No. 5,536,637), or was identified as encoding a secreted or transmembrane protein on the basis of computer analysis of the amino acid sequence of the encoded protein. fk354_4 is a full-length clone, including the entire coding sequence of a secreted protein (also referred to herein as "fk354_4 protein").

The nucleotide sequence of fk354_4 as presently determined is reported in SEQ ID NO:15. What applicants presently believe to be the proper reading frame and the predicted amino acid sequence of the fk354_4 protein corresponding to the foregoing nucleotide sequence is reported in SEQ ID NO:16.

The EcoRI/NotI restriction fragment obtainable from the deposit containing clone fk354_4 should be approximately 1800 bp.

The nucleotide sequence disclosed herein for fk354_4 was searched against the GenBank and GeneSeq nucleotide sequence databases using BLASTN/BLASTX and FASTA search protocols. fk354_4 demonstrated at least some similarity with sequences identified as AA086801 (mm85d09.r1 Stratagene mouse embryonic carcinomaRA (#937318) Mus musculus cDNA clone 535217 5' similar to SW:YE04_YEAST P32642 HYPOTHETICAL 27.5 KD PROTEIN IN RAD3-BMH1 INTERGENIC REGION), H17927 (ym41g12.s1 Homo sapiens cDNA clone 50743 3'), H78479 (yu12d02.r1 Homo sapiens cDNA clone 233571 5' similar to SP THIH_TOBAC P29449 THIOREDOXIN), W14808 (mb32g03.r1 Soares mouse p3NMF19), W49686 (zc43g10.s1 Soares senescent fibroblasts

NbHSF Homo sapiens cDNA clone 325122 3' similar to SW YE04_YEAST P32642 HYPOTHETICAL 27.5 KD PROTEIN IN RAD3-BMH1 INTERGENIC REGION), W58564 (zd19b11.r1 Soares fetal heart NbHH19W Homo sapiens cDNA clone 341085 5' similar to SW:YE04_YEAST P32642 HYPOTHETICAL 27.5 KD PROTEIN IN RAD3-BMH1 INTERGENIC REGION), and W73086 (zd54b10.r1 Soares fetal heart NbHH19W Homo sapiens cDNA clone 344443 5' similar to SW:YE04_YEAST P32642 HYPOTHETICAL 27.5 KD PROTEIN IN RAD3-BMH1 INTERGENIC REGION). The predicted amino acid sequence disclosed herein for fk354_4 was searched against the GenPept and GeneSeq amino acid sequence databases using the BLASTX search protocol. The predicted fk354_4 protein demonstrated at least some similarity to sequences identified as R50051 (ICP34.5 fragment), R93017 (Hard wheat thioredoxin h), U18922 (Yer174p [Saccharomyces cerevisiae]), and Z47746 (probable thioredoxin [Saccharomyces cerevisiae]). Based upon sequence similarity, fk354_4 proteins and each similar protein or peptide may share at least some activity.

Clone "fm150_1"

A polynucleotide of the present invention has been identified as clone "fm150_1". fm150_1 was isolated from a human adult brain cDNA library using methods which are selective for cDNAs encoding secreted proteins (see U.S. Pat. No. 5,536,637), or was identified as encoding a secreted or transmembrane protein on the basis of computer analysis of the amino acid sequence of the encoded protein. fm150_1 is a full-length clone, including the entire coding sequence of a secreted protein (also referred to herein as "fm150_1 protein").

The nucleotide sequence of fm150_1 as presently determined is reported in SEQ ID NO:17. What applicants presently believe to be the proper reading frame and the predicted amino acid sequence of the fm150_1 protein corresponding to the foregoing nucleotide sequence is reported in SEQ ID NO:18.

The EcoRI/NotI restriction fragment obtainable from the deposit containing clone fm150_1 should be approximately 1400 bp.

The nucleotide sequence disclosed herein for fm150_1 was searched against the GenBank and GeneSeq nucleotide sequence databases using BLASTN/BLASTX and FASTA search protocols. fm150_1 demonstrated at least some similarity with sequences identified as AA035409 (zk26h11.r1 Soares pregnant uterus NbHPU Homo sapiens cDNA clone 471717 5' similar to WP F22B5.2 CE02197 RNA BINDING PROTEIN), AA046762

(zk72c04.r1 Soares pregnant uterus NbHPU Homo sapiens cDNA clone 488358 5' similar to WP:F22B5.2 CE02197 RNA BINDING PROTEIN), AA135078 (zo26d06.r1 Stratagene colon (#937204) Homo sapiens cDNA clone 588011 5'), AF020833 (Homo sapiens eukaryotic translation initiation factor 3 subunit (p42) mRNA, complete cds), M78660 (EST00808 Homo sapiens cDNA clone HHCMA48), Q60681 (Human brain Expressed Sequence Tag EST00808), and Z99383 (Homo sapiens mRNA; expressed sequence tag; clone DKFZphamy1_1b5, 5' read). The predicted amino acid sequence disclosed herein for fm150_1 was searched against the GenPept and GeneSeq amino acid sequence databases using the BLASTX search protocol. The predicted fm150_1 protein demonstrated at least some similarity to sequences identified as AF004913 (translation initiation factor 3 p33 subunit; Tif35p [Saccharomyces cerevisiae]), AF020833 (eukaryotic translation initiation factor 3 subunit [Homo sapiens]), and Z50044 (F22B5.2 [Caenorhabditis elegans]). Based upon sequence similarity, fm150_1 proteins and each similar protein or peptide may share at least some activity.

Clone "gu534_1"

A polynucleotide of the present invention has been identified as clone "gu534_1". gu534_1 was isolated from a human fetal kidney cDNA library using methods which are selective for cDNAs encoding secreted proteins (see U.S. Pat. No. 5,536,637), or was identified as encoding a secreted or transmembrane protein on the basis of computer analysis of the amino acid sequence of the encoded protein. gu534_1 is a full-length clone, including the entire coding sequence of a secreted protein (also referred to herein as "gu534_1 protein").

The nucleotide sequence of gu534_1 as presently determined is reported in SEQ ID NO:19. What applicants presently believe to be the proper reading frame and the predicted amino acid sequence of the gu534_1 protein corresponding to the foregoing nucleotide sequence is reported in SEQ ID NO:20.

The EcoRI/NotI restriction fragment obtainable from the deposit containing clone gu534_1 should be approximately 1800 bp.

The nucleotide sequence disclosed herein for gu534_1 was searched against the GenBank and GeneSeq nucleotide sequence databases using BLASTN/BLASTX and FASTA search protocols. gu534_1 demonstrated at least some similarity with sequences identified as AA186601 (zp71a10.s1 Stratagene endothelial cell 937223 Homo sapiens cDNA clone 625626 3'), AA229724 (nc48c08.s1 NCI CGAP Pr3 Homo sapiens cDNA clone

5511), AA418331 (zv96a10.r1 Soares NhHMPu S1 Homo sapiens cDNA clone 767610 5'), H30057 (yp44d12.s1 Homo sapiens cDNA clone 190295 3'), N80681 (zb03c03.s1 Homo sapiens cDNA clone 300964 3'), and W19081 (zb14d11.r1 Soares fetal lung NbHL19W Homo sapiens cDNA clone 302037 5' similar to contains element THR repetitive element). Based upon sequence similarity, gu534_1 proteins and each similar protein or peptide may share at least some activity.

Deposit of Clones

Clones en539_8, eq188_1, er80_1, er418_5, fa252_8, fg912_1, fg949_3, fk354_4, fm150_1, and gu534_1 were deposited on April 15, 1997 with the American Type Culture Collection (10801 University Boulevard, Manassas, Virginia 20110-2209 U.S.A.) as an original deposit under the Budapest Treaty and were given the accession number ATCC 98408, from which each clone comprising a particular polynucleotide is obtainable. All restrictions on the availability to the public of the deposited material will be irrevocably removed upon the granting of the patent, except for the requirements specified in 37 C.F.R. § 1.808(b), and the term of the deposit will comply with 37 C.F.R. § 1.806.

Each clone has been transfected into separate bacterial cells (*E. coli*) in this composite deposit. Each clone (except for en539_8) can be removed from the vector in which it was deposited by performing an EcoRI/NotI digestion (5' site, EcoRI; 3' site, NotI) to produce the appropriate fragment for such clone. The en539_8 clone can be removed from the vector in which it was deposited by performing an EcoRI digestion, as the insert for that clone has EcoRI sites at both its 5' and 3' ends. Each clone was deposited in either the pED6 or pNOTs vector depicted in Figures 1A and 1B, respectively. The pED6dpc2 vector ("pED6") was derived from pED6dpc1 by insertion of a new polylinker to facilitate cDNA cloning (Kaufman *et al.*, 1991, *Nucleic Acids Res.* 19: 4485-4490); the pNOTs vector was derived from pMT2 (Kaufman *et al.*, 1989, *Mol. Cell. Biol.* 9: 946-958) by deletion of the DHFR sequences, insertion of a new polylinker, and insertion of the M13 origin of replication in the ClaI site. In some instances, the deposited clone can become "flipped" (i.e., in the reverse orientation) in the deposited isolate. In such instances, the cDNA insert can still be isolated by digestion with EcoRI and NotI. However, NotI will then produce the 5' site and EcoRI will produce the 3' site for placement of the cDNA in proper orientation for expression in a suitable vector. The cDNA may also be expressed from the vectors in which they were deposited.

Bacterial cells containing a particular clone can be obtained from the composite deposit as follows:

An oligonucleotide probe or probes should be designed to the sequence that is known for that particular clone. This sequence can be derived from the sequences provided herein, or from a combination of those sequences. The sequence of an oligonucleotide probe that was used to isolate or to sequence each full-length clone is identified below, and should be most reliable in isolating the clone of interest.

Clone	Probe Sequence
en539_8	SEQ ID NO:21
eq188_1	SEQ ID NO:22
er80_1	SEQ ID NO:23
er418_5	SEQ ID NO:24
fa252_8	SEQ ID NO:25
fg912_1	SEQ ID NO:26
fg949_3	SEQ ID NO:27
fk354_4	SEQ ID NO:28
fm150_1	SEQ ID NO:29
gu534_1	SEQ ID NO:30

In the sequences listed above which include an N at position 2, that position is occupied in preferred probes/primers by a biotinylated phosphoaramidite residue rather than a nucleotide (such as , for example, that produced by use of biotin phosphoramidite (1-dimethoxytrityloxy-2-(N-biotinyl-4-aminobutyl)-propyl-3-O-(2-cyanoethyl)-(N,N-diisopropyl)-phosphoramadite) (Glen Research, cat. no. 10-1953)).

The design of the oligonucleotide probe should preferably follow these parameters:

(a) It should be designed to an area of the sequence which has the fewest ambiguous bases ("N's"), if any;

(b) It should be designed to have a T_m of approx. 80 ° C (assuming 2° for each A or T and 4 degrees for each G or C).

The oligonucleotide should preferably be labeled with g-^{32}P ATP (specific activity 6000 Ci/mmole) and T4 polynucleotide kinase using commonly employed techniques for labeling oligonucleotides. Other labeling techniques can also be used. Unincorporated

label should preferably be removed by gel filtration chromatography or other established methods. The amount of radioactivity incorporated into the probe should be quantitated by measurement in a scintillation counter. Preferably, specific activity of the resulting probe should be approximately 4e+6 dpm/pmole.

The bacterial culture containing the pool of full-length clones should preferably be thawed and 100 µl of the stock used to inoculate a sterile culture flask containing 25 ml of sterile L-broth containing ampicillin at 100 µg/ml. The culture should preferably be grown to saturation at 37°C, and the saturated culture should preferably be diluted in fresh L-broth. Aliquots of these dilutions should preferably be plated to determine the dilution and volume which will yield approximately 5000 distinct and well-separated colonies on solid bacteriological media containing L-broth containing ampicillin at 100 µg/ml and agar at 1.5% in a 150 mm petri dish when grown overnight at 37°C. Other known methods of obtaining distinct, well-separated colonies can also be employed.

Standard colony hybridization procedures should then be used to transfer the colonies to nitrocellulose filters and lyse, denature and bake them.

The filter is then preferably incubated at 65°C for 1 hour with gentle agitation in 6X SSC (20X stock is 175.3 g NaCl/liter, 88.2 g Na citrate/liter, adjusted to pH 7.0 with NaOH) containing 0.5% SDS, 100 µg/ml of yeast RNA, and 10 mM EDTA (approximately 10 mL per 150 mm filter). Preferably, the probe is then added to the hybridization mix at a concentration greater than or equal to 1e+6 dpm/mL. The filter is then preferably incubated at 65°C with gentle agitation overnight. The filter is then preferably washed in 500 mL of 2X SSC/0.5% SDS at room temperature without agitation, preferably followed by 500 mL of 2X SSC/0.1% SDS at room temperature with gentle shaking for 15 minutes. A third wash with 0.1X SSC/0.5% SDS at 65°C for 30 minutes to 1 hour is optional. The filter is then preferably dried and subjected to autoradiography for sufficient time to visualize the positives on the X-ray film. Other known hybridization methods can also be employed.

The positive colonies are picked, grown in culture, and plasmid DNA isolated using standard procedures. The clones can then be verified by restriction analysis, hybridization analysis, or DNA sequencing.

Fragments of the proteins of the present invention which are capable of exhibiting biological activity are also encompassed by the present invention. Fragments of the protein may be in linear form or they may be cyclized using known methods, for example, as described in H.U. Saragovi, *et al.*, Bio/Technology 10, 773-778 (1992) and in R.S.

McDowell, *et al.*, J. Amer. Chem. Soc. 114, 9245-9253 (1992), both of which are incorporated herein by reference. Such fragments may be fused to carrier molecules such as immunoglobulins for many purposes, including increasing the valency of protein binding sites. For example, fragments of the protein may be fused through "linker" sequences to the Fc portion of an immunoglobulin. For a bivalent form of the protein, such a fusion could be to the Fc portion of an IgG molecule. Other immunoglobulin isotypes may also be used to generate such fusions. For example, a protein - IgM fusion would generate a decavalent form of the protein of the invention.

The present invention also provides both full-length and mature forms of the disclosed proteins. The full-length form of the such proteins is identified in the sequence listing by translation of the nucleotide sequence of each disclosed clone. The mature form(s) of such protein may be obtained by expression of the disclosed full-length polynucleotide (preferably those deposited with ATCC) in a suitable mammalian cell or other host cell. The sequence(s) of the mature form(s) of the protein may also be determinable from the amino acid sequence of the full-length form.

The present invention also provides genes corresponding to the polynucleotide sequences disclosed herein. "Corresponding genes" are the regions of the genome that are transcribed to produce the mRNAs from which cDNA polynucleotide sequences are derived and may include contiguous regions of the genome necessary for the regulated expression of such genes. Corresponding genes may therefore include but are not limited to coding sequences, 5' and 3' untranslated regions, alternatively spliced exons, introns, promoters, enhancers, and silencer or suppressor elements. The corresponding genes can be isolated in accordance with known methods using the sequence information disclosed herein. Such methods include the preparation of probes or primers from the disclosed sequence information for identification and/or amplification of genes in appropriate genomic libraries or other sources of genomic materials. An "isolated gene" is a gene that has been separated from the adjacent coding sequences, if any, present in the genome of the organism from which the gene was isolated.

Organisms that have enhanced, reduced, or modified expression of the gene(s) corresponding to the polynucleotide sequences disclosed herein are provided. The desired change in gene expression can be achieved through the use of antisense polynucleotides or ribozymes that bind and/or cleave the mRNA transcribed from the gene (Albert and Morris, 1994, *Trends Pharmacol. Sci.* 15(7): 250-254; Lavarosky *et al.*, 1997, *Biochem. Mol. Med.* 62(1): 11-22; and Hampel, 1998, *Prog. Nucleic Acid Res. Mol. Biol.* 58: 1-

39; all of which are incorporated by reference herein). Transgenic animals that have multiple copies of the gene(s) corresponding to the polynucleotide sequences disclosed herein, preferably produced by transformation of cells with genetic constructs that are stably maintained within the transformed cells and their progeny, are provided. Transgenic animals that have modified genetic control regions that increase or reduce gene expression levels, or that change temporal or spatial patterns of gene expression, are also provided (see European Patent No. 0 649 464 B1, incorporated by reference herein). In addition, organisms are provided in which the gene(s) corresponding to the polynucleotide sequences disclosed herein have been partially or completely inactivated, through insertion of extraneous sequences into the corresponding gene(s) or through deletion of all or part of the corresponding gene(s). Partial or complete gene inactivation can be accomplished through insertion, preferably followed by imprecise excision, of transposable elements (Plasterk, 1992, *Bioessays* **14(9)**: 629-633; Zwaal *et al.*, 1993, *Proc. Natl. Acad. Sci. USA* **90(16)**: 7431-7435; Clark *et al.*, 1994, *Proc. Natl. Acad. Sci. USA* **91(2)**: 719-722; all of which are incorporated by reference herein), or through homologous recombination, preferably detected by positive/negative genetic selection strategies (Mansour *et al.*, 1988, *Nature* 336: 348-352; U.S. Patent Nos. 5,464,764; 5,487,992; 5,627,059; 5,631,153; 5,614, 396; 5,616,491; and 5,679,523; all of which are incorporated by reference herein). These organisms with altered gene expression are preferably eukaryotes and more preferably are mammals. Such organisms are useful for the development of non-human models for the study of disorders involving the corresponding gene(s), and for the development of assay systems for the identification of molecules that interact with the protein product(s) of the corresponding gene(s).

Where the protein of the present invention is membrane-bound (e.g., is a receptor), the present invention also provides for soluble forms of such protein. In such forms part or all of the intracellular and transmembrane domains of the protein are deleted such that the protein is fully secreted from the cell in which it is expressed. The intracellular and transmembrane domains of proteins of the invention can be identified in accordance with known techniques for determination of such domains from sequence information.

Proteins and protein fragments of the present invention include proteins with amino acid sequence lengths that are at least 25%(more preferably at least 50%, and most preferably at least 75%) of the length of a disclosed protein and have at least 60% sequence identity (more preferably, at least 75% identity; most preferably at least 90% or 95% identity) with that disclosed protein, where sequence identity is determined by comparing

the amino acid sequences of the proteins when aligned so as to maximize overlap and identity while minimizing sequence gaps. Also included in the present invention are proteins and protein fragments that contain a segment preferably comprising 8 or more (more preferably 20 or more, most preferably 30 or more) contiguous amino acids that shares at least 75% sequence identity (more preferably, at least 85% identity; most preferably at least 95% identity) with any such segment of any of the disclosed proteins.

Species homologues of the disclosed polynucleotides and proteins are also provided by the present invention. As used herein, a "species homologue" is a protein or polynucleotide with a different species of origin from that of a given protein or polynucleotide, but with significant sequence similarity to the given protein or polynucleotide. Preferably, polynucleotide species homologues have at least 60% sequence identity (more preferably, at least 75% identity; most preferably at least 90% identity) with the given polynucleotide, and protein species homologues have at least 30% sequence identity (more preferably, at least 45% identity; most preferably at least 60% identity) with the given protein, where sequence identity is determined by comparing the nucleotide sequences of the polynucleotides or the amino acid sequences of the proteins when aligned so as to maximize overlap and identity while minimizing sequence gaps. Species homologues may be isolated and identified by making suitable probes or primers from the sequences provided herein and screening a suitable nucleic acid source from the desired species. Preferably, species homologues are those isolated from mammalian species. Most preferably, species homologues are those isolated from certain mammalian species such as, for example, *Pan troglodytes, Gorilla gorilla, Pongo pygmaeus, Hylobates concolor, Macaca mulatta, Papio papio, Papio hamadryas, Cercopithecus aethiops, Cebus capucinus, Aotus trivirgatus, Sanguinus oedipus, Microcebus murinus, Mus musculus, Rattus norvegicus, Cricetulus griseus, Felis catus, Mustela vison, Canis familiaris, Oryctolagus cuniculus, Bos taurus, Ovis aries, Sus scrofa,* and *Equus caballus,* for which genetic maps have been created allowing the identification of syntenic relationships between the genomic organization of genes in one species and the genomic organization of the related genes in another species (O'Brien and Seuánez, 1988, *Ann. Rev. Genet.* 22: 323-351; O'Brien *et al.*, 1993, *Nature Genetics* 3:103-112; Johansson *et al.*, 1995, *Genomics* 25: 682-690; Lyons *et al.*, 1997, *Nature Genetics* 15: 47-56; O'Brien *et al.*, 1997, *Trends in Genetics* 13(10): 393-399; Carver and Stubbs, 1997, *Genome Research* 7:1123-1137; all of which are incorporated by reference herein).

The invention also encompasses allelic variants of the disclosed polynucleotides or proteins; that is, naturally-occurring alternative forms of the isolated polynucleotides which also encode proteins which are identical or have significantly similar sequences to those encoded by the disclosed polynucleotides. Preferably, allelic variants have at least 60% sequence identity (more preferably, at least 75% identity; most preferably at least 90% identity) with the given polynucleotide, where sequence identity is determined by comparing the nucleotide sequences of the polynucleotides when aligned so as to maximize overlap and identity while minimizing sequence gaps. Allelic variants may be isolated and identified by making suitable probes or primers from the sequences provided herein and screening a suitable nucleic acid source from individuals of the appropriate species.

The invention also includes polynucleotides with sequences complementary to those of the polynucleotides disclosed herein.

The present invention also includes polynucleotides that hybridize under reduced stringency conditions, more preferably stringent conditions, and most preferably highly stringent conditions, to polynucleotides described herein. Examples of stringency conditions are shown in the table below: highly stringent conditions are those that are at least as stringent as, for example, conditions A-F; stringent conditions are at least as stringent as, for example, conditions G-L; and reduced stringency conditions are at least as stringent as, for example, conditions M-R.

Stringency Condition	Polynucleotide Hybrid	Hybrid Length (bp)‡	Hybridization Temperature and Buffer†	Wash Temperature and Buffer†
A	DNA:DNA	≥ 50	65°C; 1xSSC -or- 42°C; 1xSSC, 50% formamide	65°C; 0.3xSSC
B	DNA:DNA	<50	T_B*; 1xSSC	T_B*; 1xSSC
C	DNA:RNA	≥ 50	67°C; 1xSSC -or- 45°C; 1xSSC, 50% formamide	67°C; 0.3xSSC
D	DNA:RNA	<50	T_D*; 1xSSC	T_D*; 1xSSC
E	RNA:RNA	≥ 50	70°C; 1xSSC -or- 50°C; 1xSSC, 50% formamide	70°C; 0.3xSSC
F	RNA:RNA	<50	T_F*; 1xSSC	T_F*; 1xSSC
G	DNA:DNA	≥ 50	65°C; 4xSSC -or- 42°C; 4xSSC, 50% formamide	65°C; 1xSSC
H	DNA:DNA	<50	T_H*; 4xSSC	T_H*; 4xSSC
I	DNA:RNA	≥ 50	67°C; 4xSSC -or- 45°C; 4xSSC, 50% formamide	67°C; 1xSSC
J	DNA:RNA	<50	T_J*; 4xSSC	T_J*; 4xSSC
K	RNA:RNA	≥ 50	70°C; 4xSSC -or- 50°C; 4xSSC, 50% formamide	67°C; 1xSSC
L	RNA:RNA	<50	T_L*; 2xSSC	T_L*; 2xSSC
M	DNA:DNA	≥ 50	50°C; 4xSSC -or- 40°C; 6xSSC, 50% formamide	50°C; 2xSSC
N	DNA:DNA	<50	T_N*; 6xSSC	T_N*; 6xSSC
O	DNA:RNA	≥ 50	55°C; 4xSSC -or- 42°C; 6xSSC, 50% formamide	55°C; 2xSSC
P	DNA:RNA	<50	T_P*; 6xSSC	T_P*; 6xSSC
Q	RNA:RNA	≥ 50	60°C; 4xSSC -or- 45°C; 6xSSC, 50% formamide	60°C; 2xSSC
R	RNA:RNA	<50	T_R*; 4xSSC	T_R*; 4xSSC

‡: The hybrid length is that anticipated for the hybridized region(s) of the hybridizing polynucleotides. When hybridizing a polynucleotide to a target polynucleotide of unknown sequence, the hybrid length is assumed to be that of the hybridizing polynucleotide. When polynucleotides of known sequence are hybridized, the hybrid length can be determined by aligning the sequences of the polynucleotides and identifying the region or regions of optimal sequence complementarity.

†: SSPE (1xSSPE is 0.15M NaCl, 10mM NaH_2PO_4, and 1.25mM EDTA, pH 7.4) can be substituted for SSC (1xSSC is 0.15M NaCl and 15mM sodium citrate) in the hybridization and wash buffers; washes are performed for 15 minutes after hybridization is complete.

*T_B - T_R: The hybridization temperature for hybrids anticipated to be less than 50 base pairs in length should be 5-10°C less than the melting temperature (T_m) of the hybrid, where T_m is determined according to the following equations. For hybrids less than 18 base pairs in length, T_m(°C) = 2(# of A + T bases) + 4(# of G + C bases). For hybrids between 18 and 49 base pairs in length, T_m(°C) = 81.5 + 16.6($\log_{10}$[Na^+]) + 0.41(%G+C) - (600/N), where N is the number of bases in the hybrid, and [Na^+] is the concentration of sodium ions in the hybridization buffer ([Na^+] for 1xSSC = 0.165 M).

Additional examples of stringency conditions for polynucleotide hybridization are provided in Sambrook, J., E.F. Fritsch, and T. Maniatis, 1989, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, chapters 9 and 11, and *Current Protocols in Molecular Biology*, 1995, F.M. Ausubel et al., eds., John Wiley & Sons, Inc., sections 2.10 and 6.3-6.4, incorporated herein by reference.

Preferably, each such hybridizing polynucleotide has a length that is at least 25%(more preferably at least 50%, and most preferably at least 75%) of the length of the polynucleotide of the present invention to which it hybridizes, and has at least 60% sequence identity (more preferably, at least 75% identity; most preferably at least 90% or 95% identity) with the polynucleotide of the present invention to which it hybridizes, where sequence identity is determined by comparing the sequences of the hybridizing polynucleotides when aligned so as to maximize overlap and identity while minimizing sequence gaps.

The isolated polynucleotide of the invention may be operably linked to an expression control sequence such as the pMT2 or pED expression vectors disclosed in Kaufman *et al.*, Nucleic Acids Res. 19, 4485-4490 (1991), in order to produce the protein recombinantly. Many suitable expression control sequences are known in the art. General methods of expressing recombinant proteins are also known and are exemplified in R. Kaufman, Methods in Enzymology 185, 537-566 (1990). As defined herein "operably linked" means that the isolated polynucleotide of the invention and an expression control sequence are situated within a vector or cell in such a way that the protein is expressed by a host cell which has been transformed (transfected) with the ligated polynucleotide/expression control sequence.

A number of types of cells may act as suitable host cells for expression of the protein. Mammalian host cells include, for example, monkey COS cells, Chinese Hamster Ovary (CHO) cells, human kidney 293 cells, human epidermal A431 cells, human Colo205 cells, 3T3 cells, CV-1 cells, other transformed primate cell lines, normal diploid cells, cell strains derived from in vitro culture of primary tissue, primary explants, HeLa cells, mouse L cells, BHK, HL-60, U937, HaK or Jurkat cells.

Alternatively, it may be possible to produce the protein in lower eukaryotes such as yeast or in prokaryotes such as bacteria. Potentially suitable yeast strains include *Saccharomyces cerevisiae, Schizosaccharomyces pombe, Kluyveromyces* strains, *Candida*, or any yeast strain capable of expressing heterologous proteins. Potentially suitable bacterial strains include *Escherichia coli, Bacillus subtilis, Salmonella typhimurium*, or any bacterial

strain capable of expressing heterologous proteins. If the protein is made in yeast or bacteria, it may be necessary to modify the protein produced therein, for example by phosphorylation or glycosylation of the appropriate sites, in order to obtain the functional protein. Such covalent attachments may be accomplished using known chemical or enzymatic methods.

The protein may also be produced by operably linking the isolated polynucleotide of the invention to suitable control sequences in one or more insect expression vectors, and employing an insect expression system. Materials and methods for baculovirus/insect cell expression systems are commercially available in kit form from, *e.g.*, Invitrogen, San Diego, California, U.S.A. (the MaxBac® kit), and such methods are well known in the art, as described in Summers and Smith, Texas Agricultural Experiment Station Bulletin No. 1555 (1987), incorporated herein by reference. As used herein, an insect cell capable of expressing a polynucleotide of the present invention is "transformed."

The protein of the invention may be prepared by culturing transformed host cells under culture conditions suitable to express the recombinant protein. The resulting expressed protein may then be purified from such culture (i.e., from culture medium or cell extracts) using known purification processes, such as gel filtration and ion exchange chromatography. The purification of the protein may also include an affinity column containing agents which will bind to the protein; one or more column steps over such affinity resins as concanavalin A-agarose, heparin-toyopearl® or Cibacrom blue 3GA Sepharose®; one or more steps involving hydrophobic interaction chromatography using such resins as phenyl ether, butyl ether, or propyl ether; or immunoaffinity chromatography.

Alternatively, the protein of the invention may also be expressed in a form which will facilitate purification. For example, it may be expressed as a fusion protein, such as those of maltose binding protein (MBP), glutathione-S-transferase (GST) or thioredoxin (TRX). Kits for expression and purification of such fusion proteins are commercially available from New England BioLab (Beverly, MA), Pharmacia (Piscataway, NJ) and InVitrogen, respectively. The protein can also be tagged with an epitope and subsequently purified by using a specific antibody directed to such epitope. One such epitope ("Flag") is commercially available from Kodak (New Haven, CT).

Finally, one or more reverse-phase high performance liquid chromatography (RP-HPLC) steps employing hydrophobic RP-HPLC media, e.g., silica gel having pendant

methyl or other aliphatic groups, can be employed to further purify the protein. Some or all of the foregoing purification steps, in various combinations, can also be employed to provide a substantially homogeneous isolated recombinant protein. The protein thus purified is substantially free of other mammalian proteins and is defined in accordance with the present invention as an "isolated protein."

The protein of the invention may also be expressed as a product of transgenic animals, e.g., as a component of the milk of transgenic cows, goats, pigs, or sheep which are characterized by somatic or germ cells containing a nucleotide sequence encoding the protein.

The protein may also be produced by known conventional chemical synthesis. Methods for constructing the proteins of the present invention by synthetic means are known to those skilled in the art. The synthetically-constructed protein sequences, by virtue of sharing primary, secondary or tertiary structural and/or conformational characteristics with proteins may possess biological properties in common therewith, including protein activity. Thus, they may be employed as biologically active or immunological substitutes for natural, purified proteins in screening of therapeutic compounds and in immunological processes for the development of antibodies.

The proteins provided herein also include proteins characterized by amino acid sequences similar to those of purified proteins but into which modification are naturally provided or deliberately engineered. For example, modifications in the peptide or DNA sequences can be made by those skilled in the art using known techniques. Modifications of interest in the protein sequences may include the alteration, substitution, replacement, insertion or deletion of a selected amino acid residue in the coding sequence. For example, one or more of the cysteine residues may be deleted or replaced with another amino acid to alter the conformation of the molecule. Techniques for such alteration, substitution, replacement, insertion or deletion are well known to those skilled in the art (see, e.g., U.S. Patent No. 4,518,584). Preferably, such alteration, substitution, replacement, insertion or deletion retains the desired activity of the protein.

Other fragments and derivatives of the sequences of proteins which would be expected to retain protein activity in whole or in part and may thus be useful for screening or other immunological methodologies may also be easily made by those skilled in the art given the disclosures herein. Such modifications are believed to be encompassed by the present invention.

USES AND BIOLOGICAL ACTIVITY

The polynucleotides and proteins of the present invention are expected to exhibit one or more of the uses or biological activities (including those associated with assays cited herein) identified below. Uses or activities described for proteins of the present invention may be provided by administration or use of such proteins or by administration or use of polynucleotides encoding such proteins (such as, for example, in gene therapies or vectors suitable for introduction of DNA).

Research Uses and Utilities

The polynucleotides provided by the present invention can be used by the research community for various purposes. The polynucleotides can be used to express recombinant protein for analysis, characterization or therapeutic use; as markers for tissues in which the corresponding protein is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in disease states); as molecular weight markers on Southern gels; as chromosome markers or tags (when labeled) to identify chromosomes or to map related gene positions; to compare with endogenous DNA sequences in patients to identify potential genetic disorders; as probes to hybridize and thus discover novel, related DNA sequences; as a source of information to derive PCR primers for genetic fingerprinting; as a probe to "subtract-out" known sequences in the process of discovering other novel polynucleotides; for selecting and making oligomers for attachment to a "gene chip" or other support, including for examination of expression patterns; to raise anti-protein antibodies using DNA immunization techniques; and as an antigen to raise anti-DNA antibodies or elicit another immune response. Where the polynucleotide encodes a protein which binds or potentially binds to another protein (such as, for example, in a receptor-ligand interaction), the polynucleotide can also be used in interaction trap assays (such as, for example, those described in Gyuris *et al.*, 1993, *Cell* 75: 791-803 and in Rossi *et al.*, 1997, *Proc. Natl. Acad. Sci. USA* 94: 8405-8410, all of which are incorporated by reference herein) to identify polynucleotides encoding the other protein with which binding occurs or to identify inhibitors of the binding interaction.

The proteins provided by the present invention can similarly be used in assay to determine biological activity, including in a panel of multiple proteins for high-throughput screening; to raise antibodies or to elicit another immune response; as a reagent (including the labeled reagent) in assays designed to quantitatively determine

levels of the protein (or its receptor) in biological fluids; as markers for tissues in which the corresponding protein is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in a disease state); and, of course, to isolate correlative receptors or ligands. Where the protein binds or potentially binds to another protein (such as, for example, in a receptor-ligand interaction), the protein can be used to identify the other protein with which binding occurs or to identify inhibitors of the binding interaction. Proteins involved in these binding interactions can also be used to screen for peptide or small molecule inhibitors or agonists of the binding interaction.

Any or all of these research utilities are capable of being developed into reagent grade or kit format for commercialization as research products.

Methods for performing the uses listed above are well known to those skilled in the art. References disclosing such methods include without limitation "Molecular Cloning: A Laboratory Manual", 2d ed., Cold Spring Harbor Laboratory Press, Sambrook, J., E.F. Fritsch and T. Maniatis eds., 1989, and "Methods in Enzymology: Guide to Molecular Cloning Techniques", Academic Press, Berger, S.L. and A.R. Kimmel eds., 1987.

Nutritional Uses

Polynucleotides and proteins of the present invention can also be used as nutritional sources or supplements. Such uses include without limitation use as a protein or amino acid supplement, use as a carbon source, use as a nitrogen source and use as a source of carbohydrate. In such cases the protein or polynucleotide of the invention can be added to the feed of a particular organism or can be administered as a separate solid or liquid preparation, such as in the form of powder, pills, solutions, suspensions or capsules. In the case of microorganisms, the protein or polynucleotide of the invention can be added to the medium in or on which the microorganism is cultured.

Cytokine and Cell Proliferation/Differentiation Activity

A protein of the present invention may exhibit cytokine, cell proliferation (either inducing or inhibiting) or cell differentiation (either inducing or inhibiting) activity or may induce production of other cytokines in certain cell populations. Many protein factors discovered to date, including all known cytokines, have exhibited activity in one or more factor dependent cell proliferation assays, and hence the assays serve as a convenient confirmation of cytokine activity. The activity of a protein of the present invention is

evidenced by any one of a number of routine factor dependent cell proliferation assays for cell lines including, without limitation, 32D, DA2, DA1G, T10, B9, B9/11, BaF3, MC9/G, M+ (preB M+), 2E8, RB5, DA1, 123, T1165, HT2, CTLL2, TF-1, Mo7e and CMK.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Assays for T-cell or thymocyte proliferation include without limitation those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A.M. Kruisbeek, D.H. Margulies, E.M. Shevach, W Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, In Vitro assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Takai et al., J. Immunol. 137:3494-3500, 1986; Bertagnolli et al., J. Immunol. 145:1706-1712, 1990; Bertagnolli et al., Cellular Immunology 133:327-341, 1991; Bertagnolli, et al., J. Immunol. 149:3778-3783, 1992; Bowman et al., J. Immunol. 152: 1756-1761, 1994.

Assays for cytokine production and/or proliferation of spleen cells, lymph node cells or thymocytes include, without limitation, those described in: Polyclonal T cell stimulation, Kruisbeek, A.M. and Shevach, E.M. In *Current Protocols in Immunology*. J.E.e.a. Coligan eds. Vol 1 pp. 3.12.1-3.12.14, John Wiley and Sons, Toronto. 1994; and Measurement of mouse and human Interferon γ, Schreiber, R.D. In *Current Protocols in Immunology*. J.E.e.a. Coligan eds. Vol 1 pp. 6.8.1-6.8.8, John Wiley and Sons, Toronto. 1994.

Assays for proliferation and differentiation of hematopoietic and lymphopoietic cells include, without limitation, those described in: Measurement of Human and Murine Interleukin 2 and Interleukin 4, Bottomly, K., Davis, L.S. and Lipsky, P.E. In *Current Protocols in Immunology*. J.E.e.a. Coligan eds. Vol 1 pp. 6.3.1-6.3.12, John Wiley and Sons, Toronto. 1991; deVries et al., J. Exp. Med. 173:1205-1211, 1991; Moreau et al., Nature 336:690-692, 1988; Greenberger et al., Proc. Natl. Acad. Sci. U.S.A. 80:2931-2938, 1983; Measurement of mouse and human interleukin 6 - Nordan, R. In *Current Protocols in Immunology*. J.E.e.a. Coligan eds. Vol 1 pp. 6.6.1-6.6.5, John Wiley and Sons, Toronto. 1991; Smith et al., Proc. Natl. Acad. Sci. U.S.A. 83:1857-1861, 1986; Measurement of human Interleukin 11 - Bennett, F., Giannotti, J., Clark, S.C. and Turner, K. J. In *Current Protocols in Immunology*. J.E.e.a. Coligan eds. Vol 1 pp. 6.15.1 John Wiley and Sons, Toronto. 1991; Measurement of mouse and human Interleukin 9 - Ciarletta, A., Giannotti, J., Clark, S.C. and Turner, K.J. In *Current Protocols in Immunology*. J.E.e.a. Coligan eds. Vol 1 pp. 6.13.1, John Wiley and Sons, Toronto. 1991.

Assays for T-cell clone responses to antigens (which will identify, among others, proteins that affect APC-T cell interactions as well as direct T-cell effects by measuring proliferation and cytokine production) include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A.M. Kruisbeek, D.H. Margulies, E.M. Shevach, W Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, In Vitro assays for Mouse Lymphocyte Function; Chapter 6, Cytokines and their cellular receptors; Chapter 7, Immunologic studies in Humans); Weinberger et al., Proc. Natl. Acad. Sci. USA 77:6091-6095, 1980; Weinberger et al., Eur. J. Immun. 11:405-411, 1981; Takai et al., J. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988.

Immune Stimulating or Suppressing Activity

A protein of the present invention may also exhibit immune stimulating or immune suppressing activity, including without limitation the activities for which assays are described herein. A protein may be useful in the treatment of various immune deficiencies and disorders (including severe combined immunodeficiency (SCID)), e.g., in regulating (up or down) growth and proliferation of T and/or B lymphocytes, as well as effecting the cytolytic activity of NK cells and other cell populations. These immune deficiencies may be genetic or be caused by viral (e.g., HIV) as well as bacterial or fungal infections, or may result from autoimmune disorders. More specifically, infectious diseases causes by viral, bacterial, fungal or other infection may be treatable using a protein of the present invention, including infections by HIV, hepatitis viruses, herpesviruses, mycobacteria, Leishmania spp., malaria spp. and various fungal infections such as candidiasis. Of course, in this regard, a protein of the present invention may also be useful where a boost to the immune system generally may be desirable, *i.e.*, in the treatment of cancer.

Autoimmune disorders which may be treated using a protein of the present invention include, for example, connective tissue disease, multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation, Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent diabetes mellitis, myasthenia gravis, graft-versus-host disease and autoimmune inflammatory eye disease. Such a protein of the present invention may also to be useful in the treatment of allergic reactions and conditions, such as asthma (particularly allergic asthma) or other respiratory problems. Other conditions, in which immune suppression is desired (including, for

example, organ transplantation), may also be treatable using a protein of the present invention.

Using the proteins of the invention it may also be possible to immune responses, in a number of ways. Down regulation may be in the form of inhibiting or blocking an immune response already in progress or may involve preventing the induction of an immune response. The functions of activated T cells may be inhibited by suppressing T cell responses or by inducing specific tolerance in T cells, or both. Immunosuppression of T cell responses is generally an active, non-antigen-specific, process which requires continuous exposure of the T cells to the suppressive agent. Tolerance, which involves inducing non-responsiveness or anergy in T cells, is distinguishable from immunosuppression in that it is generally antigen-specific and persists after exposure to the tolerizing agent has ceased. Operationally, tolerance can be demonstrated by the lack of a T cell response upon reexposure to specific antigen in the absence of the tolerizing agent.

Down regulating or preventing one or more antigen functions (including without limitation B lymphocyte antigen functions (such as , for example, B7)), *e.g.*, preventing high level lymphokine synthesis by activated T cells, will be useful in situations of tissue, skin and organ transplantation and in graft-versus-host disease (GVHD). For example, blockage of T cell function should result in reduced tissue destruction in tissue transplantation. Typically, in tissue transplants, rejection of the transplant is initiated through its recognition as foreign by T cells, followed by an immune reaction that destroys the transplant. The administration of a molecule which inhibits or blocks interaction of a B7 lymphocyte antigen with its natural ligand(s) on immune cells (such as a soluble, monomeric form of a peptide having B7-2 activity alone or in conjunction with a monomeric form of a peptide having an activity of another B lymphocyte antigen (*e.g.*, B7-1, B7-3) or blocking antibody), prior to transplantation can lead to the binding of the molecule to the natural ligand(s) on the immune cells without transmitting the corresponding costimulatory signal. Blocking B lymphocyte antigen function in this matter prevents cytokine synthesis by immune cells, such as T cells, and thus acts as an immunosuppressant. Moreover, the lack of costimulation may also be sufficient to anergize the T cells, thereby inducing tolerance in a subject. Induction of long-term tolerance by B lymphocyte antigen-blocking reagents may avoid the necessity of repeated administration of these blocking reagents. To achieve sufficient immunosuppression or

tolerance in a subject, it may also be necessary to block the function of a combination of B lymphocyte antigens.

The efficacy of particular blocking reagents in preventing organ transplant rejection or GVHD can be assessed using animal models that are predictive of efficacy in humans. Examples of appropriate systems which can be used include allogeneic cardiac grafts in rats and xenogeneic pancreatic islet cell grafts in mice, both of which have been used to examine the immunosuppressive effects of CTLA4Ig fusion proteins *in vivo* as described in Lenschow *et al.*, Science *257*:789-792 (1992) and Turka *et al.*, Proc. Natl. Acad. Sci USA, *89*:11102-11105 (1992). In addition, murine models of GVHD (see Paul ed., Fundamental Immunology, Raven Press, New York, 1989, pp. 846-847) can be used to determine the effect of blocking B lymphocyte antigen function *in vivo* on the development of that disease.

Blocking antigen function may also be therapeutically useful for treating autoimmune diseases. Many autoimmune disorders are the result of inappropriate activation of T cells that are reactive against self tissue and which promote the production of cytokines and autoantibodies involved in the pathology of the diseases. Preventing the activation of autoreactive T cells may reduce or eliminate disease symptoms. Administration of reagents which block costimulation of T cells by disrupting receptor:ligand interactions of B lymphocyte antigens can be used to inhibit T cell activation and prevent production of autoantibodies or T cell-derived cytokines which may be involved in the disease process. Additionally, blocking reagents may induce antigen-specific tolerance of autoreactive T cells which could lead to long-term relief from the disease. The efficacy of blocking reagents in preventing or alleviating autoimmune disorders can be determined using a number of well-characterized animal models of human autoimmune diseases. Examples include murine experimental autoimmune encephalitis, systemic lupus erythmatosis in MRL/*lpr/lpr* mice or NZB hybrid mice, murine autoimmune collagen arthritis, diabetes mellitus in NOD mice and BB rats, and murine experimental myasthenia gravis (see Paul ed., Fundamental Immunology, Raven Press, New York, 1989, pp. 840-856).

Upregulation of an antigen function (preferably a B lymphocyte antigen function), as a means of up regulating immune responses, may also be useful in therapy. Upregulation of immune responses may be in the form of enhancing an existing immune response or eliciting an initial immune response. For example, enhancing an immune response through stimulating B lymphocyte antigen function may be useful in cases of

viral infection. In addition, systemic viral diseases such as influenza, the common cold, and encephalitis might be alleviated by the administration of stimulatory forms of B lymphocyte antigens systemically.

Alternatively, anti-viral immune responses may be enhanced in an infected patient by removing T cells from the patient, costimulating the T cells *in vitro* with viral antigen-pulsed APCs either expressing a peptide of the present invention or together with a stimulatory form of a soluble peptide of the present invention and reintroducing the *in vitro* activated T cells into the patient. Another method of enhancing anti-viral immune responses would be to isolate infected cells from a patient, transfect them with a nucleic acid encoding a protein of the present invention as described herein such that the cells express all or a portion of the protein on their surface, and reintroduce the transfected cells into the patient. The infected cells would now be capable of delivering a costimulatory signal to, and thereby activate, T cells *in vivo*.

In another application, up regulation or enhancement of antigen function (preferably B lymphocyte antigen function) may be useful in the induction of tumor immunity. Tumor cells (*e.g.*, sarcoma, melanoma, lymphoma, leukemia, neuroblastoma, carcinoma) transfected with a nucleic acid encoding at least one peptide of the present invention can be administered to a subject to overcome tumor-specific tolerance in the subject. If desired, the tumor cell can be transfected to express a combination of peptides. For example, tumor cells obtained from a patient can be transfected *ex vivo* with an expression vector directing the expression of a peptide having B7-2-like activity alone, or in conjunction with a peptide having B7-1-like activity and/or B7-3-like activity. The transfected tumor cells are returned to the patient to result in expression of the peptides on the surface of the transfected cell. Alternatively, gene therapy techniques can be used to target a tumor cell for transfection *in vivo*.

The presence of the peptide of the present invention having the activity of a B lymphocyte antigen(s) on the surface of the tumor cell provides the necessary costimulation signal to T cells to induce a T cell mediated immune response against the transfected tumor cells. In addition, tumor cells which lack MHC class I or MHC class II molecules, or which fail to reexpress sufficient amounts of MHC class I or MHC class II molecules, can be transfected with nucleic acid encoding all or a portion of (*e.g.*, a cytoplasmic-domain truncated portion) of an MHC class I α chain protein and β_2 microglobulin protein or an MHC class II α chain protein and an MHC class II β chain protein to thereby express MHC class I or MHC class II proteins on the cell surface.

Expression of the appropriate class I or class II MHC in conjunction with a peptide having the activity of a B lymphocyte antigen (*e.g.*, B7-1, B7-2, B7-3) induces a T cell mediated immune response against the transfected tumor cell. Optionally, a gene encoding an antisense construct which blocks expression of an MHC class II associated protein, such as the invariant chain, can also be cotransfected with a DNA encoding a peptide having the activity of a B lymphocyte antigen to promote presentation of tumor associated antigens and induce tumor specific immunity. Thus, the induction of a T cell mediated immune response in a human subject may be sufficient to overcome tumor-specific tolerance in the subject.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Suitable assays for thymocyte or splenocyte cytotoxicity include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A.M. Kruisbeek, D.H. Margulies, E.M. Shevach, W Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, In Vitro assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Herrmann et al., Proc. Natl. Acad. Sci. USA 78:2488-2492, 1981; Herrmann et al., J. Immunol. 128:1968-1974, 1982; Handa et al., J. Immunol. 135:1564-1572, 1985; Takai et al., J. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988; Herrmann et al., Proc. Natl. Acad. Sci. USA 78:2488-2492, 1981; Herrmann et al., J. Immunol. 128:1968-1974, 1982; Handa et al., J. Immunol. 135:1564-1572, 1985; Takai et al., J. Immunol. 137:3494-3500, 1986; Bowmanet al., J. Virology 61:1992-1998; Takai et al., J. Immunol. 140:508-512, 1988; Bertagnolli et al., Cellular Immunology 133:327-341, 1991; Brown et al., J. Immunol. 153:3079-3092, 1994.

Assays for T-cell-dependent immunoglobulin responses and isotype switching (which will identify, among others, proteins that modulate T-cell dependent antibody responses and that affect Th1/Th2 profiles) include, without limitation, those described in: Maliszewski, J. Immunol. 144:3028-3033, 1990; and Assays for B cell function: *In vitro* antibody production, Mond, J.J. and Brunswick, M. In *Current Protocols in Immunology*. J.E.e.a. Coligan eds. Vol 1 pp. 3.8.1-3.8.16, John Wiley and Sons, Toronto. 1994.

Mixed lymphocyte reaction (MLR) assays (which will identify, among others, proteins that generate predominantly Th1 and CTL responses) include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A.M. Kruisbeek, D.H. Margulies, E.M. Shevach, W Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, In Vitro assays for Mouse Lymphocyte Function 3.1-3.19; Chapter

7, Immunologic studies in Humans); Takai et al., J. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988; Bertagnolli et al., J. Immunol. 149:3778-3783, 1992.

Dendritic cell-dependent assays (which will identify, among others, proteins expressed by dendritic cells that activate naive T-cells) include, without limitation, those described in: Guery et al., J. Immunol. 134:536-544, 1995; Inaba et al., Journal of Experimental Medicine 173:549-559, 1991; Macatonia et al., Journal of Immunology 154:5071-5079, 1995; Porgador et al., Journal of Experimental Medicine 182:255-260, 1995; Nair et al., Journal of Virology 67:4062-4069, 1993; Huang et al., Science 264:961-965, 1994; Macatonia et al., Journal of Experimental Medicine 169:1255-1264, 1989; Bhardwaj et al., Journal of Clinical Investigation 94:797-807, 1994; and Inaba et al., Journal of Experimental Medicine 172:631-640, 1990.

Assays for lymphocyte survival/apoptosis (which will identify, among others, proteins that prevent apoptosis after superantigen induction and proteins that regulate lymphocyte homeostasis) include, without limitation, those described in: Darzynkiewicz et al., Cytometry 13:795-808, 1992; Gorczyca et al., Leukemia 7:659-670, 1993; Gorczyca et al., Cancer Research 53:1945-1951, 1993; Itoh et al., Cell 66:233-243, 1991; Zacharchuk, Journal of Immunology 145:4037-4045, 1990; Zamai et al., Cytometry 14:891-897, 1993; Gorczyca et al., International Journal of Oncology 1:639-648, 1992.

Assays for proteins that influence early steps of T-cell commitment and development include, without limitation, those described in: Antica et al., Blood 84:111-117, 1994; Fine et al., Cellular Immunology 155:111-122, 1994; Galy et al., Blood 85:2770-2778, 1995; Toki et al., Proc. Nat. Acad Sci. USA 88:7548-7551, 1991.

Hematopoiesis Regulating Activity

A protein of the present invention may be useful in regulation of hematopoiesis and, consequently, in the treatment of myeloid or lymphoid cell deficiencies. Even marginal biological activity in support of colony forming cells or of factor-dependent cell lines indicates involvement in regulating hematopoiesis, e.g. in supporting the growth and proliferation of erythroid progenitor cells alone or in combination with other cytokines, thereby indicating utility, for example, in treating various anemias or for use in conjunction with irradiation/chemotherapy to stimulate the production of erythroid precursors and/or erythroid cells; in supporting the growth and proliferation of myeloid cells such as granulocytes and monocytes/macrophages (i.e., traditional CSF activity) useful, for example, in conjunction with chemotherapy to prevent or treat consequent

myelo-suppression; in supporting the growth and proliferation of megakaryocytes and consequently of platelets thereby allowing prevention or treatment of various platelet disorders such as thrombocytopenia, and generally for use in place of or complimentary to platelet transfusions; and/or in supporting the growth and proliferation of hematopoietic stem cells which are capable of maturing to any and all of the above-mentioned hematopoietic cells and therefore find therapeutic utility in various stem cell disorders (such as those usually treated with transplantation, including, without limitation, aplastic anemia and paroxysmal nocturnal hemoglobinuria), as well as in repopulating the stem cell compartment post irradiation/chemotherapy, either *in-vivo* or *ex-vivo* (i.e., in conjunction with bone marrow transplantation or with peripheral progenitor cell transplantation (homologous or heterologous)) as normal cells or genetically manipulated for gene therapy.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Suitable assays for proliferation and differentiation of various hematopoietic lines are cited above.

Assays for embryonic stem cell differentiation (which will identify, among others, proteins that influence embryonic differentiation hematopoiesis) include, without limitation, those described in: Johansson et al. Cellular Biology 15:141-151, 1995; Keller et al., Molecular and Cellular Biology 13:473-486, 1993; McClanahan et al., Blood 81:2903-2915, 1993.

Assays for stem cell survival and differentiation (which will identify, among others, proteins that regulate lympho-hematopoiesis) include, without limitation, those described in: Methylcellulose colony forming assays, Freshney, M.G. In *Culture of Hematopoietic Cells*. R.I. Freshney, *et al.* eds. Vol pp. 265-268, Wiley-Liss, Inc., New York, NY. 1994; Hirayama et al., Proc. Natl. Acad. Sci. USA 89:5907-5911, 1992; Primitive hematopoietic colony forming cells with high proliferative potential, McNiece, I.K. and Briddell, R.A. In *Culture of Hematopoietic Cells*. R.I. Freshney, *et al.* eds. Vol pp. 23-39, Wiley-Liss, Inc., New York, NY. 1994; Neben et al., Experimental Hematology 22:353-359, 1994; Cobblestone area forming cell assay, Ploemacher, R.E. In *Culture of Hematopoietic Cells*. R.I. Freshney, *et al.* eds. Vol pp. 1-21, Wiley-Liss, Inc.., New York, NY. 1994; Long term bone marrow cultures in the presence of stromal cells, Spooncer, E., Dexter, M. and Allen, T. In *Culture of Hematopoietic Cells*. R.I. Freshney, *et al.* eds. Vol pp. 163-179, Wiley-Liss, Inc., New York, NY. 1994; Long term culture initiating cell assay, Sutherland,

H.J. In *Culture of Hematopoietic Cells*. R.I. Freshney, *et al.* eds. Vol pp. 139-162, Wiley-Liss, Inc., New York, NY. 1994.

Tissue Growth Activity

A protein of the present invention also may have utility in compositions used for bone, cartilage, tendon, ligament and/or nerve tissue growth or regeneration, as well as for wound healing and tissue repair and replacement, and in the treatment of burns, incisions and ulcers.

A protein of the present invention, which induces cartilage and/or bone growth in circumstances where bone is not normally formed, has application in the healing of bone fractures and cartilage damage or defects in humans and other animals. Such a preparation employing a protein of the invention may have prophylactic use in closed as well as open fracture reduction and also in the improved fixation of artificial joints. *De novo* bone formation induced by an osteogenic agent contributes to the repair of congenital, trauma induced, or oncologic resection induced craniofacial defects, and also is useful in cosmetic plastic surgery.

A protein of this invention may also be used in the treatment of periodontal disease, and in other tooth repair processes. Such agents may provide an environment to attract bone-forming cells, stimulate growth of bone-forming cells or induce differentiation of progenitors of bone-forming cells. A protein of the invention may also be useful in the treatment of osteoporosis or osteoarthritis, such as through stimulation of bone and/or cartilage repair or by blocking inflammation or processes of tissue destruction (collagenase activity, osteoclast activity, etc.) mediated by inflammatory processes.

Another category of tissue regeneration activity that may be attributable to the protein of the present invention is tendon/ligament formation. A protein of the present invention, which induces tendon/ligament-like tissue or other tissue formation in circumstances where such tissue is not normally formed, has application in the healing of tendon or ligament tears, deformities and other tendon or ligament defects in humans and other animals. Such a preparation employing a tendon/ligament-like tissue inducing protein may have prophylactic use in preventing damage to tendon or ligament tissue, as well as use in the improved fixation of tendon or ligament to bone or other tissues, and in repairing defects to tendon or ligament tissue. De novo tendon/ligament-like tissue formation induced by a composition of the present invention contributes to the repair of

congenital, trauma induced, or other tendon or ligament defects of other origin, and is also useful in cosmetic plastic surgery for attachment or repair of tendons or ligaments. The compositions of the present invention may provide an environment to attract tendon- or ligament-forming cells, stimulate growth of tendon- or ligament-forming cells, induce differentiation of progenitors of tendon- or ligament-forming cells, or induce growth of tendon/ligament cells or progenitors *ex vivo* for return *in vivo* to effect tissue repair. The compositions of the invention may also be useful in the treatment of tendinitis, carpal tunnel syndrome and other tendon or ligament defects. The compositions may also include an appropriate matrix and/or sequestering agent as a carrier as is well known in the art.

The protein of the present invention may also be useful for proliferation of neural cells and for regeneration of nerve and brain tissue, *i.e.* for the treatment of central and peripheral nervous system diseases and neuropathies, as well as mechanical and traumatic disorders, which involve degeneration, death or trauma to neural cells or nerve tissue. More specifically, a protein may be used in the treatment of diseases of the peripheral nervous system, such as peripheral nerve injuries, peripheral neuropathy and localized neuropathies, and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome. Further conditions which may be treated in accordance with the present invention include mechanical and traumatic disorders, such as spinal cord disorders, head trauma and cerebrovascular diseases such as stroke. Peripheral neuropathies resulting from chemotherapy or other medical therapies may also be treatable using a protein of the invention.

Proteins of the invention may also be useful to promote better or faster closure of non-healing wounds, including without limitation pressure ulcers, ulcers associated with vascular insufficiency, surgical and traumatic wounds, and the like.

It is expected that a protein of the present invention may also exhibit activity for generation or regeneration of other tissues, such as organs (including, for example, pancreas, liver, intestine, kidney, skin, endothelium), muscle (smooth, skeletal or cardiac) and vascular (including vascular endothelium) tissue, or for promoting the growth of cells comprising such tissues. Part of the desired effects may be by inhibition or modulation of fibrotic scarring to allow normal tissue to regenerate. A protein of the invention may also exhibit angiogenic activity.

A protein of the present invention may also be useful for gut protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues, and conditions resulting from systemic cytokine damage.

A protein of the present invention may also be useful for promoting or inhibiting differentiation of tissues described above from precursor tissues or cells; or for inhibiting the growth of tissues described above.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Assays for tissue generation activity include, without limitation, those described in: International Patent Publication No. WO95/16035 (bone, cartilage, tendon); International Patent Publication No. WO95/05846 (nerve, neuronal); International Patent Publication No. WO91/07491 (skin, endothelium).

Assays for wound healing activity include, without limitation, those described in: Winter, Epidermal Wound Healing, pps. 71-112 (Maibach, HI and Rovee, DT, eds.), Year Book Medical Publishers, Inc., Chicago, as modified by Eaglstein and Mertz, J. Invest. Dermatol 71:382-84 (1978).

Activin/Inhibin Activity

A protein of the present invention may also exhibit activin- or inhibin-related activities. Inhibins are characterized by their ability to inhibit the release of follicle stimulating hormone (FSH), while activins and are characterized by their ability to stimulate the release of follicle stimulating hormone (FSH). Thus, a protein of the present invention, alone or in heterodimers with a member of the inhibin α family, may be useful as a contraceptive based on the ability of inhibins to decrease fertility in female mammals and decrease spermatogenesis in male mammals. Administration of sufficient amounts of other inhibins can induce infertility in these mammals. Alternatively, the protein of the invention, as a homodimer or as a heterodimer with other protein subunits of the inhibin-β group, may be useful as a fertility inducing therapeutic, based upon the ability of activin molecules in stimulating FSH release from cells of the anterior pituitary. See, for example, United States Patent 4,798,885. A protein of the invention may also be useful for advancement of the onset of fertility in sexually immature mammals, so as to increase the lifetime reproductive performance of domestic animals such as cows, sheep and pigs.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Assays for activin/inhibin activity include, without limitation, those described in: Vale et al., Endocrinology 91:562-572, 1972; Ling et al., Nature 321:779-782, 1986; Vale et al., Nature 321:776-779, 1986; Mason et al., Nature 318:659-663, 1985; Forage et al., Proc. Natl. Acad. Sci. USA 83:3091-3095, 1986.

Chemotactic/Chemokinetic Activity

A protein of the present invention may have chemotactic or chemokinetic activity (e.g., act as a chemokine) for mammalian cells, including, for example, monocytes, fibroblasts, neutrophils, T-cells, mast cells, eosinophils, epithelial and/or endothelial cells. Chemotactic and chemokinetic proteins can be used to mobilize or attract a desired cell population to a desired site of action. Chemotactic or chemokinetic proteins provide particular advantages in treatment of wounds and other trauma to tissues, as well as in treatment of localized infections. For example, attraction of lymphocytes, monocytes or neutrophils to tumors or sites of infection may result in improved immune responses against the tumor or infecting agent.

A protein or peptide has chemotactic activity for a particular cell population if it can stimulate, directly or indirectly, the directed orientation or movement of such cell population. Preferably, the protein or peptide has the ability to directly stimulate directed movement of cells. Whether a particular protein has chemotactic activity for a population of cells can be readily determined by employing such protein or peptide in any known assay for cell chemotaxis.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Assays for chemotactic activity (which will identify proteins that induce or prevent chemotaxis) consist of assays that measure the ability of a protein to induce the migration of cells across a membrane as well as the ability of a protein to induce the adhesion of one cell population to another cell population. Suitable assays for movement and adhesion include, without limitation, those described in: Current Protocols in Immunology, Ed by J.E. Coligan, A.M. Kruisbeek, D.H. Margulies, E.M. Shevach, W.Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 6.12, Measurement of alpha and beta Chemokines 6.12.1-6.12.28; Taub et al. J. Clin. Invest. 95:1370-1376, 1995; Lind et al. APMIS 103:140-146, 1995; Muller et al Eur. J. Immunol. 25: 1744-1748; Gruber et al. J. of Immunol. 152:5860-5867, 1994; Johnston et al. J. of Immunol. 153: 1762-1768, 1994.

Hemostatic and Thrombolytic Activity

A protein of the invention may also exhibit hemostatic or thrombolytic activity. As a result, such a protein is expected to be useful in treatment of various coagulation disorders (including hereditary disorders, such as hemophilias) or to enhance coagulation and other hemostatic events in treating wounds resulting from trauma, surgery or other causes. A protein of the invention may also be useful for dissolving or inhibiting formation of thromboses and for treatment and prevention of conditions resulting therefrom (such as, for example, infarction of cardiac and central nervous system vessels (e.g., stroke).

The activity of a protein of the invention may, among other means, be measured by the following methods:

Assay for hemostatic and thrombolytic activity include, without limitation, those described in: Linet et al., J. Clin. Pharmacol. 26:131-140, 1986; Burdick et al., Thrombosis Res. 45:413-419, 1987; Humphrey et al., Fibrinolysis 5:71-79 (1991); Schaub, Prostaglandins 35:467-474, 1988.

Receptor/Ligand Activity

A protein of the present invention may also demonstrate activity as receptors, receptor ligands or inhibitors or agonists of receptor/ligand interactions. Examples of such receptors and ligands include, without limitation, cytokine receptors and their ligands, receptor kinases and their ligands, receptor phosphatases and their ligands, receptors involved in cell-cell interactions and their ligands (including without limitation, cellular adhesion molecules (such as selectins, integrins and their ligands) and receptor/ligand pairs involved in antigen presentation, antigen recognition and development of cellular and humoral immune responses). Receptors and ligands are also useful for screening of potential peptide or small molecule inhibitors of the relevant receptor/ligand interaction. A protein of the present invention (including, without limitation, fragments of receptors and ligands) may themselves be useful as inhibitors of receptor/ligand interactions.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Suitable assays for receptor-ligand activity include without limitation those described in:Current Protocols in Immunology, Ed by J.E. Coligan, A.M. Kruisbeek, D.H. Margulies, E.M. Shevach, W.Strober, Pub. Greene Publishing Associates and

Wiley-Interscience (Chapter 7.28, Measurement of Cellular Adhesion under static conditions 7.28.1-7.28.22), Takai et al., Proc. Natl. Acad. Sci. USA 84:6864-6868, 1987; Bierer et al., J. Exp. Med. 168:1145-1156, 1988; Rosenstein et al., J. Exp. Med. 169:149-160 1989; Stoltenborg et al., J. Immunol. Methods 175:59-68, 1994; Stitt et al., Cell 80:661-670, 1995.

Anti-Inflammatory Activity

Proteins of the present invention may also exhibit anti-inflammatory activity. The anti-inflammatory activity may be achieved by providing a stimulus to cells involved in the inflammatory response, by inhibiting or promoting cell-cell interactions (such as, for example, cell adhesion), by inhibiting or promoting chemotaxis of cells involved in the inflammatory process, inhibiting or promoting cell extravasation, or by stimulating or suppressing production of other factors which more directly inhibit or promote an inflammatory response. Proteins exhibiting such activities can be used to treat inflammatory conditions including chronic or acute conditions), including without limitation inflammation associated with infection (such as septic shock, sepsis or systemic inflammatory response syndrome (SIRS)), ischemia-reperfusion injury, endotoxin lethality, arthritis, complement-mediated hyperacute rejection, nephritis, cytokine or chemokine-induced lung injury, inflammatory bowel disease, Crohn's disease or resulting from over production of cytokines such as TNF or IL-1. Proteins of the invention may also be useful to treat anaphylaxis and hypersensitivity to an antigenic substance or material.

Cadherin/Tumor Invasion Suppressor Activity

Cadherins are calcium-dependent adhesion molecules that appear to play major roles during development, particularly in defining specific cell types. Loss or alteration of normal cadherin expression can lead to changes in cell adhesion properties linked to tumor growth and metastasis. Cadherin malfunction is also implicated in other human diseases, such as pemphigus vulgaris and pemphigus foliaceus (auto-immune blistering skin diseases), Crohn's disease, and some developmental abnormalities.

The cadherin superfamily includes well over forty members, each with a distinct pattern of expression. All members of the superfamily have in common conserved extracellular repeats (cadherin domains), but structural differences are found in other parts of the molecule. The cadherin domains bind calcium to form their tertiary structure and thus calcium is required to mediate their adhesion. Only a few amino acids in the

first cadherin domain provide the basis for homophilic adhesion; modification of this recognition site can change the specificity of a cadherin so that instead of recognizing only itself, the mutant molecule can now also bind to a different cadherin. In addition, some cadherins engage in heterophilic adhesion with other cadherins.

E-cadherin, one member of the cadherin superfamily, is expressed in epithelial cell types. Pathologically, if E-cadherin expression is lost in a tumor, the malignant cells become invasive and the cancer metastasizes. Transfection of cancer cell lines with polynucleotides expressing E-cadherin has reversed cancer-associated changes by returning altered cell shapes to normal, restoring cells' adhesiveness to each other and to their substrate, decreasing the cell growth rate, and drastically reducing anchorage-independent cell growth. Thus, reintroducing E-cadherin expression reverts carcinomas to a less advanced stage. It is likely that other cadherins have the same invasion suppressor role in carcinomas derived from other tissue types. Therefore, proteins of the present invention with cadherin activity, and polynucleotides of the present invention encoding such proteins, can be used to treat cancer. Introducing such proteins or polynucleotides into cancer cells can reduce or eliminate the cancerous changes observed in these cells by providing normal cadherin expression.

Cancer cells have also been shown to express cadherins of a different tissue type than their origin, thus allowing these cells to invade and metastasize in a different tissue in the body. Proteins of the present invention with cadherin activity, and polynucleotides of the present invention encoding such proteins, can be substituted in these cells for the inappropriately expressed cadherins, restoring normal cell adhesive properties and reducing or eliminating the tendency of the cells to metastasize.

Additionally, proteins of the present invention with cadherin activity, and polynucleotides of the present invention encoding such proteins, can used to generate antibodies recognizing and binding to cadherins. Such antibodies can be used to block the adhesion of inappropriately expressed tumor-cell cadherins, preventing the cells from forming a tumor elsewhere. Such an anti-cadherin antibody can also be used as a marker for the grade, pathological type, and prognosis of a cancer, i.e. the more progressed the cancer, the less cadherin expression there will be, and this decrease in cadherin expression can be detected by the use of a cadherin-binding antibody.

Fragments of proteins of the present invention with cadherin activity, preferably a polypeptide comprising a decapeptide of the cadherin recognition site, and polynucleotides of the present invention encoding such protein fragments, can also be used

to block cadherin function by binding to cadherins and preventing them from binding in ways that produce undesirable effects. Additionally, fragments of proteins of the present invention with cadherin activity, preferably truncated soluble cadherin fragments which have been found to be stable in the circulation of cancer patients, and polynucleotides encoding such protein fragments, can be used to disturb proper cell-cell adhesion.

Assays for cadherin adhesive and invasive suppressor activity include, without limitation, those described in: Hortsch et al. J Biol Chem 270 (32): 18809-18817, 1995; Miyaki et al. Oncogene 11: 2547-2552, 1995; Ozawa et al. Cell 63: 1033-1038, 1990.

Tumor Inhibition Activity

In addition to the activities described above for immunological treatment or prevention of tumors, a protein of the invention may exhibit other anti-tumor activities. A protein may inhibit tumor growth directly or indirectly (such as, for example, via ADCC). A protein may exhibit its tumor inhibitory activity by acting on tumor tissue or tumor precursor tissue, by inhibiting formation of tissues necessary to support tumor growth (such as, for example, by inhibiting angiogenesis), by causing production of other factors, agents or cell types which inhibit tumor growth, or by suppressing, eliminating or inhibiting factors, agents or cell types which promote tumor growth.

Other Activities

A protein of the invention may also exhibit one or more of the following additional activities or effects: inhibiting the growth, infection or function of, or killing, infectious agents, including, without limitation, bacteria, viruses, fungi and other parasites; effecting (suppressing or enhancing) bodily characteristics, including, without limitation, height, weight, hair color, eye color, skin, fat to lean ratio or other tissue pigmentation, or organ or body part size or shape (such as, for example, breast augmentation or diminution, change in bone form or shape); effecting biorhythms or caricadic cycles or rhythms; effecting the fertility of male or female subjects; effecting the metabolism, catabolism, anabolism, processing, utilization, storage or elimination of dietary fat, lipid, protein, carbohydrate, vitamins, minerals, cofactors or other nutritional factors or component(s); effecting behavioral characteristics, including, without limitation, appetite, libido, stress, cognition (including cognitive disorders), depression (including depressive disorders) and violent behaviors; providing analgesic effects or other pain reducing effects; promoting differentiation and growth of embryonic stem cells in lineages other than hematopoietic

lineages; hormonal or endocrine activity; in the case of enzymes, correcting deficiencies of the enzyme and treating deficiency-related diseases; treatment of hyperproliferative disorders (such as, for example, psoriasis); immunoglobulin-like activity (such as, for example, the ability to bind antigens or complement); and the ability to act as an antigen in a vaccine composition to raise an immune response against such protein or another material or entity which is cross-reactive with such protein.

ADMINISTRATION AND DOSING

A protein of the present invention (from whatever source derived, including without limitation from recombinant and non-recombinant sources) may be used in a pharmaceutical composition when combined with a pharmaceutically acceptable carrier. Such a composition may also contain (in addition to protein and a carrier) diluents, fillers, salts, buffers, stabilizers, solubilizers, and other materials well known in the art. The term "pharmaceutically acceptable" means a non-toxic material that does not interfere with the effectiveness of the biological activity of the active ingredient(s). The characteristics of the carrier will depend on the route of administration. The pharmaceutical composition of the invention may also contain cytokines, lymphokines, or other hematopoietic factors such as M-CSF, GM-CSF, TNF, IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, IL-14, IL-15, IFN, TNF0, TNF1, TNF2, G-CSF, Meg-CSF, thrombopoietin, stem cell factor, and erythropoietin. The pharmaceutical composition may further contain other agents which either enhance the activity of the protein or compliment its activity or use in treatment. Such additional factors and/or agents may be included in the pharmaceutical composition to produce a synergistic effect with protein of the invention, or to minimize side effects. Conversely, protein of the present invention may be included in formulations of the particular cytokine, lymphokine, other hematopoietic factor, thrombolytic or anti-thrombotic factor, or anti-inflammatory agent to minimize side effects of the cytokine, lymphokine, other hematopoietic factor, thrombolytic or anti-thrombotic factor, or anti-inflammatory agent.

A protein of the present invention may be active in multimers (e.g., heterodimers or homodimers) or complexes with itself or other proteins. As a result, pharmaceutical compositions of the invention may comprise a protein of the invention in such multimeric or complexed form.

The pharmaceutical composition of the invention may be in the form of a complex of the protein(s) of present invention along with protein or peptide antigens. The protein and/or peptide antigen will deliver a stimulatory signal to both B and T lymphocytes. B lymphocytes will respond to antigen through their surface immunoglobulin receptor. T lymphocytes will respond to antigen through the T cell receptor (TCR) following presentation of the antigen by MHC proteins. MHC and structurally related proteins including those encoded by class I and class II MHC genes on host cells will serve to present the peptide antigen(s) to T lymphocytes. The antigen components could also be supplied as purified MHC-peptide complexes alone or with co-stimulatory molecules that can directly signal T cells. Alternatively antibodies able to bind surface immunolgobulin and other molecules on B cells as well as antibodies able to bind the TCR and other molecules on T cells can be combined with the pharmaceutical composition of the invention.

The pharmaceutical composition of the invention may be in the form of a liposome in which protein of the present invention is combined, in addition to other pharmaceutically acceptable carriers, with amphipathic agents such as lipids which exist in aggregated form as micelles, insoluble monolayers, liquid crystals, or lamellar layers in aqueous solution. Suitable lipids for liposomal formulation include, without limitation, monoglycerides, diglycerides, sulfatides, lysolecithin, phospholipids, saponin, bile acids, and the like. Preparation of such liposomal formulations is within the level of skill in the art, as disclosed, for example, in U.S. Patent No. 4,235,871; U.S. Patent No. 4,501,728; U.S. Patent No. 4,837,028; and U.S. Patent No. 4,737,323, all of which are incorporated herein by reference.

As used herein, the term "therapeutically effective amount" means the total amount of each active component of the pharmaceutical composition or method that is sufficient to show a meaningful patient benefit, i.e., treatment, healing, prevention or amelioration of the relevant medical condition, or an increase in rate of treatment, healing, prevention or amelioration of such conditions. When applied to an individual active ingredient, administered alone, the term refers to that ingredient alone. When applied to a combination, the term refers to combined amounts of the active ingredients that result in the therapeutic effect, whether administered in combination, serially or simultaneously.

In practicing the method of treatment or use of the present invention, a therapeutically effective amount of protein of the present invention is administered to a mammal having a condition to be treated. Protein of the present invention may be

administered in accordance with the method of the invention either alone or in combination with other therapies such as treatments employing cytokines, lymphokines or other hematopoietic factors. When co-administered with one or more cytokines, lymphokines or other hematopoietic factors, protein of the present invention may be administered either simultaneously with the cytokine(s), lymphokine(s), other hematopoietic factor(s), thrombolytic or anti-thrombotic factors, or sequentially. If administered sequentially, the attending physician will decide on the appropriate sequence of administering protein of the present invention in combination with cytokine(s), lymphokine(s), other hematopoietic factor(s), thrombolytic or anti-thrombotic factors.

Administration of protein of the present invention used in the pharmaceutical composition or to practice the method of the present invention can be carried out in a variety of conventional ways, such as oral ingestion, inhalation, topical application or cutaneous, subcutaneous, intraperitoneal, parenteral or intravenous injection. Intravenous administration to the patient is preferred.

When a therapeutically effective amount of protein of the present invention is administered orally, protein of the present invention will be in the form of a tablet, capsule, powder, solution or elixir. When administered in tablet form, the pharmaceutical composition of the invention may additionally contain a solid carrier such as a gelatin or an adjuvant. The tablet, capsule, and powder contain from about 5 to 95% protein of the present invention, and preferably from about 25 to 90% protein of the present invention. When administered in liquid form, a liquid carrier such as water, petroleum, oils of animal or plant origin such as peanut oil, mineral oil, soybean oil, or sesame oil, or synthetic oils may be added. The liquid form of the pharmaceutical composition may further contain physiological saline solution, dextrose or other saccharide solution, or glycols such as ethylene glycol, propylene glycol or polyethylene glycol. When administered in liquid form, the pharmaceutical composition contains from about 0.5 to 90% by weight of protein of the present invention, and preferably from about 1 to 50% protein of the present invention.

When a therapeutically effective amount of protein of the present invention is administered by intravenous, cutaneous or subcutaneous injection, protein of the present invention will be in the form of a pyrogen-free, parenterally acceptable aqueous solution. The preparation of such parenterally acceptable protein solutions, having due regard to pH, isotonicity, stability, and the like, is within the skill in the art. A preferred

pharmaceutical composition for intravenous, cutaneous, or subcutaneous injection should contain, in addition to protein of the present invention, an isotonic vehicle such as Sodium Chloride Injection, Ringer's Injection, Dextrose Injection, Dextrose and Sodium Chloride Injection, Lactated Ringer's Injection, or other vehicle as known in the art. The pharmaceutical composition of the present invention may also contain stabilizers, preservatives, buffers, antioxidants, or other additives known to those of skill in the art.

The amount of protein of the present invention in the pharmaceutical composition of the present invention will depend upon the nature and severity of the condition being treated, and on the nature of prior treatments which the patient has undergone. Ultimately, the attending physician will decide the amount of protein of the present invention with which to treat each individual patient. Initially, the attending physician will administer low doses of protein of the present invention and observe the patient's response. Larger doses of protein of the present invention may be administered until the optimal therapeutic effect is obtained for the patient, and at that point the dosage is not increased further. It is contemplated that the various pharmaceutical compositions used to practice the method of the present invention should contain about 0.01 µg to about 100 mg (preferably about 0.1ng to about 10 mg, more preferably about 0.1 µg to about 1 mg) of protein of the present invention per kg body weight.

The duration of intravenous therapy using the pharmaceutical composition of the present invention will vary, depending on the severity of the disease being treated and the condition and potential idiosyncratic response of each individual patient. It is contemplated that the duration of each application of the protein of the present invention will be in the range of 12 to 24 hours of continuous intravenous administration. Ultimately the attending physician will decide on the appropriate duration of intravenous therapy using the pharmaceutical composition of the present invention.

Protein of the invention may also be used to immunize animals to obtain polyclonal and monoclonal antibodies which specifically react with the protein. Such antibodies may be obtained using either the entire protein or fragments thereof as an immunogen. The peptide immunogens additionally may contain a cysteine residue at the carboxyl terminus, and are conjugated to a hapten such as keyhole limpet hemocyanin (KLH). Methods for synthesizing such peptides are known in the art, for example, as in R.P. Merrifield, J. Amer.Chem.Soc. 85, 2149-2154 (1963); J.L. Krstenansky, *et al.*, FEBS Lett. 211, 10 (1987). Monoclonal antibodies binding to the protein of the invention may be useful diagnostic agents for the immunodetection of the protein. Neutralizing monoclonal

antibodies binding to the protein may also be useful therapeutics for both conditions associated with the protein and also in the treatment of some forms of cancer where abnormal expression of the protein is involved. In the case of cancerous cells or leukemic cells, neutralizing monoclonal antibodies against the protein may be useful in detecting and preventing the metastatic spread of the cancerous cells, which may be mediated by the protein.

For compositions of the present invention which are useful for bone, cartilage, tendon or ligament regeneration, the therapeutic method includes administering the composition topically, systematically, or locally as an implant or device. When administered, the therapeutic composition for use in this invention is, of course, in a pyrogen-free, physiologically acceptable form. Further, the composition may desirably be encapsulated or injected in a viscous form for delivery to the site of bone, cartilage or tissue damage. Topical administration may be suitable for wound healing and tissue repair. Therapeutically useful agents other than a protein of the invention which may also optionally be included in the composition as described above, may alternatively or additionally, be administered simultaneously or sequentially with the composition in the methods of the invention. Preferably for bone and/or cartilage formation, the composition would include a matrix capable of delivering the protein-containing composition to the site of bone and/or cartilage damage, providing a structure for the developing bone and cartilage and optimally capable of being resorbed into the body. Such matrices may be formed of materials presently in use for other implanted medical applications.

The choice of matrix material is based on biocompatibility, biodegradability, mechanical properties, cosmetic appearance and interface properties. The particular application of the compositions will define the appropriate formulation. Potential matrices for the compositions may be biodegradable and chemically defined calcium sulfate, tricalciumphosphate, hydroxyapatite, polylactic acid, polyglycolic acid and polyanhydrides. Other potential materials are biodegradable and biologically well-defined, such as bone or dermal collagen. Further matrices are comprised of pure proteins or extracellular matrix components. Other potential matrices are nonbiodegradable and chemically defined, such as sintered hydroxapatite, bioglass, aluminates, or other ceramics. Matrices may be comprised of combinations of any of the above mentioned types of material, such as polylactic acid and hydroxyapatite or collagen and tricalciumphosphate. The bioceramics may be altered in composition, such as in calcium-

aluminate-phosphate and processing to alter pore size, particle size, particle shape, and biodegradability.

Presently preferred is a 50:50 (mole weight) copolymer of lactic acid and glycolic acid in the form of porous particles having diameters ranging from 150 to 800 microns. In some applications, it will be useful to utilize a sequestering agent, such as carboxymethyl cellulose or autologous blood clót, to prevent the protein compositions from disassociating from the matrix.

A preferred family of sequestering agents is cellulosic materials such as alkylcelluloses (including hydroxyalkylcelluloses), including methylcellulose, ethylcellulose, hydroxyethylcellulose, hydroxypropylcellulose, hydroxypropyl-methylcellulose, and carboxymethylcellulose, the most preferred being cationic salts of carboxymethylcellulose (CMC). Other preferred sequestering agents include hyaluronic acid, sodium alginate, poly(ethylene glycol), polyoxyethylene oxide, carboxyvinyl polymer and poly(vinyl alcohol). The amount of sequestering agent useful herein is 0.5-20 wt%, preferably 1-10 wt% based on total formulation weight, which represents the amount necessary to prevent desorbtion of the protein from the polymer matrix and to provide appropriate handling of the composition, yet not so much that the progenitor cells are prevented from infiltrating the matrix, thereby providing the protein the opportunity to assist the osteogenic activity of the progenitor cells.

In further compositions, proteins of the invention may be combined with other agents beneficial to the treatment of the bone and/or cartilage defect, wound, or tissue in question. These agents include various growth factors such as epidermal growth factor (EGF), platelet derived growth factor (PDGF), transforming growth factors (TGF-α and TGF-β), and insulin-like growth factor (IGF).

The therapeutic compositions are also presently valuable for veterinary applications. Particularly domestic animals and thoroughbred horses, in addition to humans, are desired patients for such treatment with proteins of the present invention.

The dosage regimen of a protein-containing pharmaceutical composition to be used in tissue regeneration will be determined by the attending physician considering various factors which modify the action of the proteins, e.g., amount of tissue weight desired to be formed, the site of damage, the condition of the damaged tissue, the size of a wound, type of damaged tissue (e.g., bone), the patient's age, sex, and diet, the severity of any infection, time of administration and other clinical factors. The dosage may vary with the type of matrix used in the reconstitution and with inclusion of other proteins in

the pharmaceutical composition. For example, the addition of other known growth factors, such as IGF I (insulin like growth factor I), to the final composition, may also effect the dosage. Progress can be monitored by periodic assessment of tissue/bone growth and/or repair, for example, X-rays, histomorphometric determinations and tetracycline labeling.

Polynucleotides of the present invention can also be used for gene therapy. Such polynucleotides can be introduced either *in vivo* or *ex vivo* into cells for expression in a mammalian subject. Polynucleotides of the invention may also be administered by other known methods for introduction of nucleic acid into a cell or organism (including, without limitation, in the form of viral vectors or naked DNA).

Cells may also be cultured *ex vivo* in the presence of proteins of the present invention in order to proliferate or to produce a desired effect on or activity in such cells. Treated cells can then be introduced *in vivo* for therapeutic purposes.

Patent and literature references cited herein are incorporated by reference as if fully set forth.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Jacobs, Kenneth
McCoy, John M.
LaVallie, Edward R.
Racie, Lisa A.
Merberg, David
Treacy, Maurice
Spaulding, Vikki
Agostino, Michael J.

(ii) TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM

(iii) NUMBER OF SEQUENCES: 30

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Genetics Institute, Inc.
(B) STREET: 87 CambridgePark Drive
(C) CITY: Cambridge
(D) STATE: MA
(E) COUNTRY: U.S.A.
(F) ZIP: 02140

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Sprunger, Suzanne A.
(B) REGISTRATION NUMBER: 41,323

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (617) 498-8284
(B) TELEFAX: (617) 876-5851

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2754 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CAGGTGGTCC TCCACCTGCC TTGGCTTCCT AAAGTGCTGG GATTACAGGC ATGAGTCACT 60

CTGCTGGCCT ATGTTCTGTT TTTGTTTTTG TTTTTGTTTT GAGACAGAGT TTCACTCTTG 120

TTGCCCAGGC TGGAGTGCAA TGGCATAATC TCGGCTCACT GCAGCCTCTG CCTCCCAGGT 180

TCAAGTGATT CTCCTGCCTC AGCCTCCTGA GTAGCTGGGA TTACAGGCAT GTGCCACCTC 240

ACCTGGCTAA TTTTGTATTT TTAGTAGAGA TGGGGTTTCT CCATGTTAGT CAGGCTGGTC 300

TTGAACTCCT GACCTCAGGT GATCTGCCCT CCTCAGCCTC CTAAAGTGCT GGGATTACAG 360

GTGTGAGCCA CTGTGCCCAG CCTTGTTTTT TGTTTTTTTG TTTTTGTTTT TTTTTTTGAC 420

AGTAGCCATC CTAATAGATA CTAAGTGGTA TCTCATTGTG GTTTTGATTG CATGCGTTCT 480

TTTTGGCTTG TTTTTTGAGA CAAGGTCTCA CTCCATCACC CAGACTGGAG CGCAGTGGTG 540

TGATCACGGC TCGTTGCAAC CTGACCCTCT TGAGCTCAGG TGATCCTCCC ACTTCACCCT 600

CCCGAGTATC TTGGAGTACA GGTGTGTGCC TGGCTGATTT TTCGTATTTT TTGTAGAGAT 660

GGGGTTTCAC CGTGTTGCTC AGGCTGCTCT CAAACTGCTG GGCTCAAACG ATCCTCCTGC 720

CTTGGCCTCC CAAAGTGCTG GGGTTACAAG CATGAACCAT TATGCCCGGC CTGCATGCAC 780

TCTTACACAC GTTTTATCTG TTACATATCC CAAGATGTGT AGTTCTTTGG GAAGCAGGAA 840

GAAATGGGGG TAACATTGAG AAGTTAAGGA AAACTGGTAT AAATTATTGG CAGCAGCTCC 900

TGATTATAGG TTTTGAGGCC TGAGTCCATG GGCAGAGTCC CTCTCCTGCA GTTCATGAGA 960

TTTGTACCCT CCAGTGACAG TACTGGGAAG GAGGGAATGC TACGTTCCAA CTCTTAGTCT 1020

TCACTTAATT TTATGACTCA AAATTCCAGC TAGATATATA GGTTACTTTT ACTGTTGGAT 1080

CACTCTGGCC CACGAATGTA TCCTGCTAAC TTGATGTGTG CTCTAACTAC CTCCTAAGTT 1140

TGGTGACAGT CGGCAGAGTT TGTGAACCAT GTGATTCCCA ACTTAAGTTA CTAACATTTT 1200

TTTTTTTTTT TTTTGAGACA GGATCTTGCT CTGTCACCCA GGCTGGAGTG CAGTGGTACG 1260

ATCTCAGCTC ACTGTAGCCT TAACCCCACC AGGCTTATGT GCTCCTCCCA CCTCAGCCTC 1320

CCGAGTAGTT GGAACTATAG GTGCATACCA CCATGCCTGG CTAATTTTTG TATTTTTTGT 1380

AGAGGCAGGG TTTTGCCCTG TTGCCCAGGC TGGTCTTGAA CTCCTGAGCT CAAGCAATCC 1440

TCCCACCTCA GCCTCCCAAA GGGTTGGGAT TACAGGTGTG AGCCACTGCA CCCGGCCAAG 1500

TTACTAACAT TTTAAGTCTA AAGTAAAAGA TTGCTTCTGT ATGTTCTCCC CCAGGTGTGT 1560

AGGTCCATCC TGGGAAGGCC ATCAGACACA CCTAGTCCAT GGGTGACACC CAGCCAGTTT 1620

TTAATGCCAG TTCCTCTGGC AGTTTTTAAT TTAGGCACTC GGAAGTGAAA CCCGGACATT 1680

CACTGGAAAT GACTTTAGGA CAAGACCTGC TGGCCATGAG CTGAGAAATG TCTTACTCTC 1740

TTGCAGGGAG AATGCTGTTG AAAGACTTGA TTCATTAATA CAAGCGACTC ACGTTGCAAT 1800

GAGAGGCAAC TCCGATTACG CTGATCTTAG TGATGGCTGG CTCGAAATAA TACGTGTAGA 1860

TGCCCCTGAT CCAGGTGCAG ACCCGCTGGC TAGCAGTGTG AACGGCATGT GCCTGGATAT 1920

TCCTGCTCAC CTGAGCATCC GCATCCTCAT CTCGGATGCT GGCGCGGTGG AAGGGATTAC 1980

TCAGCAGGAG ATACTCGGTG TAGAGACAAG GTTCTCCTCA GTGAACTGGC AGTACCAGTG 2040

TGGGCTTACC TGTGAGCACA AGGCCGACCT TCTCCCTATC AGTGCATCCG TCCAGTTTAT 2100

TAAAATTCCT GCACAGTTAC CCCACCCCCT GACAAGATTC CAGATCAATT ATACAGAGTA 2160

TGACTGCAAC AGAAATGAGG TGTGTTGGCC GCAGCTTCTA TATCCATGGA CTCAGTATTA 2220

TCAAGGGGAG CTGCATTCTC AGTGTGTTGC TAAGGGCTTA CTGTTGCTGT TGTTCCTCAC 2280

ATTGGCCTTG TTCCTCAGCA ACCCCTGGAC CAGAATATGC AAAGCCTATA GTTAGACAAC 2340

CACCTGGCTT TTATTTTTTT GAGATGGAGT TTTGCTCTTG TTACCCAGGC TGGAGTGCAG 2400

TGCACAATCT CGGCTCACTG CAATCTCTGC CTCCCAAGCA ATCCTCCCAC CTCAGCCTCT 2460

GGTGTAGCTG GGACCACAGA TGCTCCACCA TGCCTGGCTG TATTTTTGGT AAAGATGGGG 2520

TTTCGCCTTG TTGCCCAGGG TGGTCTGTAA CTCCTGAGCT CAGATGATCT GCCCACCTCG 2580

GCCTCCCAAA GTGCTGGGAT CACAGACGTG AGCCACTGCG TCCGGTCCAT CTGACTTCTC 2640

AAAGACTTTA GACCTTGACT TCAGTGATTT GTTGTAGTCT TGTATGCTTC TCTATAAAAT 2700

TTTAATAAAT GAAATGTCTT ATTTTTGTAG AAAATTTTTA AAAAAAAAAA AAAA 2754

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 178 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Arg Gly Asn Ser Asp Tyr Ala Asp Leu Ser Asp Gly Trp Leu Glu
1 5 10 15

Ile Ile Arg Val Asp Ala Pro Asp Pro Gly Ala Asp Pro Leu Ala Ser
20 25 30

Ser Val Asn Gly Met Cys Leu Asp Ile Pro Ala His Leu Ser Ile Arg
35 40 45

Ile Leu Ile Ser Asp Ala Gly Ala Val Glu Gly Ile Thr Gln Gln Glu
50 55 60

Ile Leu Gly Val Glu Thr Arg Phe Ser Ser Val Asn Trp Gln Tyr Gln
65 70 75 80

Cys Gly Leu Thr Cys Glu His Lys Ala Asp Leu Leu Pro Ile Ser Ala
85 90 95

Ser Val Gln Phe Ile Lys Ile Pro Ala Gln Leu Pro His Pro Leu Thr
100 105 110

Arg Phe Gln Ile Asn Tyr Thr Glu Tyr Asp Cys Asn Arg Asn Glu Val
115 120 125

Cys Trp Pro Gln Leu Leu Tyr Pro Trp Thr Gln Tyr Tyr Gln Gly Glu
130 135 140

Leu His Ser Gln Cys Val Ala Lys Gly Leu Leu Leu Leu Leu Phe Leu
145 150 155 160

Thr Leu Ala Leu Phe Leu Ser Asn Pro Trp Thr Arg Ile Cys Lys Ala
165 170 175

Tyr Ser

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1363 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

TAGGCCATGA AGGCCGGTTT TTCATAAAAT AGGAATGAGG ACAAATGTTG CTCTTCATCC 60

TACCAGCTGT TTGTTCTTTG GTAGGGGATC ATGAGTGGAA AAACAAAGGC AAGAAGGGCT 120

GCCATGTTTT TTAGACGTTG CTCTGAAGAC GCCAGCGGTA GCGCCAGTGG CAATGCTTTG 180

TTATCAGAGG ACGAAAATCC TGATGCGAAT GGGGTAACTC GATCATGGAA GATTATTCTA 240

AGTACAATGC TTACACTGAC TTTTCTTCTT GTAGGACTCC TAAATCATCA GTGGCTTAAA 300

GAAACAGATG TTCCTCAGAA ATCCAGACAA TTATATGCCA TAATTGCAGA ATATGGTTCA 360

AGGCTTTATA AATATCAGGC CAGACTTCGT ATGCCTAAAG AGCAACTGGA ACTTTTAAAG 420

AAGGAAAGCC AGAATCTGGA AAACAATTTT CGTCAAATTC TATTTTTGAT CGAACAAATA 480

GATGTCCTGA AGGCATTGCT AAGAGATATG AAGGATGGTA TGGACAATAA TCACAACTGG 540

AACACCCATG GAGACCCTGT GGAGGACCCG GACCACACAG AGGAAGTGTC AAACTTGGTC 600

AATTATGTAC TTAAAAAGTT GAGAGAAGAC CAAGTCGAGA TGGCTGATTA TGCCCTGAAG 660

TCGGCCGGAG CCTCCATCAT TGAAGCTGGG ACCTCAGAAA GTTATAAAAA TAATAAAGCA 720

AAATTGTACT GGCATGGGAT AGGTTTCCTA AATCATGAAA TGCCTCCAGA TATTATTCTT 780

CAGCCGGATG TCTACCCTGG AAAGTGCTGG GCTTTTCCAG GTTCCCAGGG TCATACCCTA 840

ATCAAGCTTT ACAAAGATCA TACCAACTGC TGTTACCATG GAGCACATCT CAGAGAAGGT 900

GTCTCCGTCA GGAAACATCT CCAGTGCACC CAAGGAATTT TCTGTCTATG GCATCACAAA 960

AAAATGTGAA GGAGAAGAAA TTTTCCTAGG TCAGTTTATA TATAACAAAA CAGGAACCAC 1020

CGTTCAAACA TTTGAACTCC AGCATGCAGT TTCTGAATAT TTATTATGTG TGAAACTTAA 1080

TATCTTTAGC AACTGGGGAC ACCCGAAGTA TACTTGTTTA TATCGATTCA GGGTCCATGG 1140

CACACCAGGC AAGCACATCT AGAAGAGTTG GTACAGAAGG CCATGCCACA TGTCCAGAAT 1200

ATTCAAGAAT GCTTATTCTC TTAGATGATA CCGCACCCAT AGGAATTGAG AATTGGGAGT 1260

GGGAAGAAAA CCTCAAAGTG GTTCATACTT GCCTGTAAAA AGTAAATGCA TTTTACTAAT 1320

AAAAAAATAT GGAAGTAAAT TAAAAAAAAA AAAAAAAAAA AAA 1363

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 292 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
Met Ser Gly Lys Thr Lys Ala Arg Arg Ala Ala Met Phe Phe Arg Arg
1               5                  10                  15

Cys Ser Glu Asp Ala Ser Gly Ser Ala Ser Gly Asn Ala Leu Leu Ser
            20                  25                  30

Glu Asp Glu Asn Pro Asp Ala Asn Gly Val Thr Arg Ser Trp Lys Ile
        35                  40                  45

Ile Leu Ser Thr Met Leu Thr Leu Thr Phe Leu Leu Val Gly Leu Leu
    50                  55                  60

Asn His Gln Trp Leu Lys Glu Thr Asp Val Pro Gln Lys Ser Arg Gln
65                  70                  75                  80

Leu Tyr Ala Ile Ile Ala Glu Tyr Gly Ser Arg Leu Tyr Lys Tyr Gln
                85                  90                  95

Ala Arg Leu Arg Met Pro Lys Glu Gln Leu Glu Leu Leu Lys Lys Glu
            100                 105                 110

Ser Gln Asn Leu Glu Asn Asn Phe Arg Gln Ile Leu Phe Leu Ile Glu
        115                 120                 125

Gln Ile Asp Val Leu Lys Ala Leu Leu Arg Asp Met Lys Asp Gly Met
    130                 135                 140

Asp Asn Asn His Asn Trp Asn Thr His Gly Asp Pro Val Glu Asp Pro
145                 150                 155                 160

Asp His Thr Glu Glu Val Ser Asn Leu Val Asn Tyr Val Leu Lys Lys
                165                 170                 175

Leu Arg Glu Asp Gln Val Glu Met Ala Asp Tyr Ala Leu Lys Ser Ala
            180                 185                 190

Gly Ala Ser Ile Ile Glu Ala Gly Thr Ser Glu Ser Tyr Lys Asn Asn
        195                 200                 205

Lys Ala Lys Leu Tyr Trp His Gly Ile Gly Phe Leu Asn His Glu Met
    210                 215                 220

Pro Pro Asp Ile Ile Leu Gln Pro Asp Val Tyr Pro Gly Lys Cys Trp
225                 230                 235                 240

Ala Phe Pro Gly Ser Gln Gly His Thr Leu Ile Lys Leu Tyr Lys Asp
                245                 250                 255

His Thr Asn Cys Cys Tyr His Gly Ala His Leu Arg Glu Gly Val Ser
            260                 265                 270

Val Arg Lys His Leu Gln Cys Thr Gln Gly Ile Phe Cys Leu Trp His
```

275 280 285

His Lys Lys Met
290

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2911 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

GGGCTGCATT TCCAGCAGGA GCTGCGAGCA CAGTGCTGGC TCACAACAAG ATGCTCAAGG 60

TGTCAGCCGT ACTGTGTGTG TGTGCAGCCG CTTGGTGCAG TCAGTCTCTC GCAGCTGCCG 120

CGGCGGTGGC TGCAGCCGGG GGGCGGTCGG ACGGCGGTAA TTTTCTGGAT GATAAACAAT 180

GGCTCACCAC AATCTCTCAG TATGACAAGG AAGTCGGACA GTGGAACAAA TTCCGAGACG 240

AAGTAGAGGA TGATTATTTC CGCACTTGGA GTCCAGGAAA ACCCTTCGAT CAGGCTTTAG 300

ATCCAGCTAA GGATCCATGC TTAAAGATGA AATGTAGTCG CCATAAAGTA TGCATTGCTC 360

AAGATTCTCA GACTGCAGTC TGCATTAGTC ACCGGAGGCT TACACACAGG ATGAAAGAAG 420

CAGGAGTAGA CCATAGGCAG TGGAGGGGTC CCATATTATC CACCTGCAAG CAGTGCCCAG 480

TGGTCTATCC CAGCCCTGTT TGTGGTTCAG ATGGTCATAC CTACTCTTTT CAGTGCAAAC 540

TAGAATATCA GGCATGTGTC TTAGGAAAAC AGATCTCAGT CAAATGTGAA GGACATTGCC 600

CATGTCCTTC AGATAAGCCC ACCAGTACAA GCAGAAATGT TAAGAGAGCA TGCAGTGACC 660

TGGAGTTCAG GGAAGTGGCA AACAGATTGC GGGACTGGTT CAAGGCCCTT CATGAAAGTG 720

GAAGTCAAAA CAAGAAGACA AAAACATTGC TGAGGCCTGA GAGAAGCAGA TTCGATACCA 780

GCATCTTGCC AATTTGCAAG GACTCACTTG GCTGGATGTT TAACAGACTT GATACAAACT 840

ATGACCTGCT ATTGGACCAG TCAGAGCTCA GAAGCATTTA CCTTGATAAG AATGAACAGT 900

GTACCAAGGC ATTCTTCAAT TCTTGTGACA CATACAAGGA CAGTTTAATA TCTAATAATG 960

AGTGGTGCTA CTGCTTCCAG AGACAGCAAG ACCCACCTTG CCAGACTGAG CTCAGCAATA 1020

TTCAGAAGCG GCAAGGGGTT AAGAAGCTCC TAGGACAGTA TATCCCCCTG TGTGATGAAG 1080

ATGGTTACTA CAAGCCAACA CAATGTCATG GCAGTGTTGG ACAGTGCTGG TGTGTTGACA 1140

GATATGGAAA TGAAGTCATG GGATCCAGAA TAAATGGTGT TGCAGATTGT GCTATAGATT 1200

TTGAGATCTC CGGAGATTTT GCTAGTGGCG ATTTTCATGA ATGGACTGAT GATGAGGATG 1260

ATGAAGACGA TATTATGAAT GATGAAGATG AAATTGAAGA TGATGATGAA GATGAAGGGG 1320

ATGATGATGA TGGTGGTGAT GACCATGATG TATACATTTA ATTGATGACA GTTGAAATCA 1380

ATAAATTCTA CATTTCTAAT ATTTACAAAA ATGATAGCCT ATTTAAAATT ATCTTCTTCC 1440

CCAATAACAA AATGATTCTA AACCTCACAT ATATTTTGTA TAATTATTTG AAAAATTGCA 1500

GCTAAAGTTA TAGAACTTTA TGTTTAAATA AGAATCATTT GCTTTGAGTT TTTATATTCC 1560

TTACACAAAA AGAAAATACA TATGCAGTCT AGTCAGACAA AATAAAGTTT TGAAGTGCTA 1620

CTATAATAAG TTTTTCACGA GAACAAACTT TGTAAATCTT CCATAAGCAA AATGACAGCT 1680

AGTGCTTGGG ATCGTACATG TTAATTTTCT GAAAGATAAT TCTAAGTGAA ATTTAAAATA 1740

AATAAATTTT TAATGACCTG GGTCTTAAGG ATTTAGGAAA AATATGCATG CTTTAATTGC 1800

ATTTCCAAAG TAGCATCTTG CTAGACCTAG TTGAGTCAGG ATAACAGAGA GATACCACAT 1860

GGCAAGAAAA ACAAAGTGAC AATTGTAGAG TCCTCAATTG TGTTTACATT AATAGTGGTG 1920

TTTTTACCTA TGAAATTATT CTGGATCTAA TAGGACATTT TACAAAATGG CAAGTATGGA 1980

AAACCATGGA TTCTGAAAGT TAAAAATTTA GTTGTTCTCC CCAATGTGTA TTTTAATTTG 2040

GATGGCAGTC TCATGCAGAT TTTTTAAAAG ATTCTTTAAT AACATGATTT GTTTGCCTTT 2100

CTAGATTTCT TTATCTTTCT GACCAGCAAC TTAGGGAGCA GAATTTAAAT TAGGAAGACA 2160

AAGGGAAAGA TTCATTTAAA CCATATTTTT ACAAAGTTTG TCATTTGCCC CAAGGTCAAA 2220

TTTTAAATTC TTAATTTTCA TTTTATTTCC CATTTTAGGT AAAAGTTTGC ATTTAATCTT 2280

AGAATTATGT TATTTTTGTT AGTAGTGTGG AAACTTAGAG AACTTATTGT ATGGTGCCTT 2340

GCAAAAATAG AGATAGAAAG ATTTTAGCAT GCATACCAAT ATAGTATATT ACGCAATATA 2400

TAAGCACACC TAATTAACAG ATTAATATCA GTAAAGGTAT TGCTGCTGGA ATGAAGAAAA 2460

TGGGATACGT TTGTTTCTTT TTTTCTATTG TWACATAATT GCCATGTGGA CTTGTTTATG 2520

ATTATTGTGT AGAGTAGCAT TTAAGATTTA ACTGTAGCAA AAATTACTTT AACCGCTGTA 2580

TTTAAGTTAG CATGTTAATT AATTGTGTAG ACATTTTGGC ACACCATCAC TTTTAACTAT 2640

ATCATACCAA TGGTTTTGTG CCCATAATAA AAATGGAAAA ACCTGTTGAA TGTTACGTAT 2700

TGGTATCTTT AATTTCAACA GTGGGTAAAC TGGTTTCCCA GTATACAATT CATTGAAAGC 2760

AAAATTGATT AATTATTTCC ATTTAATTTA TACACACTCA ATACAAAATT TAATGTTGAC 2820

TTTACGTAAT AAAGTATAAT GCATTTTCTT TTTTACTGTT TATGTATAGT TTACAAAATA 2880

AAGAATCTTG TAACCAAAAA AAAAAAAAAA A 2911

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 436 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Met Leu Lys Val Ser Ala Val Leu Cys Val Cys Ala Ala Ala Trp Cys
1 5 10 15

Ser Gln Ser Leu Ala Ala Ala Ala Ala Val Ala Ala Ala Gly Gly Arg
20 25 30

Ser Asp Gly Gly Asn Phe Leu Asp Asp Lys Gln Trp Leu Thr Thr Ile
35 40 45

Ser Gln Tyr Asp Lys Glu Val Gly Gln Trp Asn Lys Phe Arg Asp Glu
50 55 60

Val Glu Asp Asp Tyr Phe Arg Thr Trp Ser Pro Gly Lys Pro Phe Asp
65 70 75 80

Gln Ala Leu Asp Pro Ala Lys Asp Pro Cys Leu Lys Met Lys Cys Ser
85 90 95

Arg His Lys Val Cys Ile Ala Gln Asp Ser Gln Thr Ala Val Cys Ile
100 105 110

Ser His Arg Arg Leu Thr His Arg Met Lys Glu Ala Gly Val Asp His
115 120 125

Arg Gln Trp Arg Gly Pro Ile Leu Ser Thr Cys Lys Gln Cys Pro Val
130 135 140

Val Tyr Pro Ser Pro Val Cys Gly Ser Asp Gly His Thr Tyr Ser Phe
145 150 155 160

Gln Cys Lys Leu Glu Tyr Gln Ala Cys Val Leu Gly Lys Gln Ile Ser
165 170 175

Val Lys Cys Glu Gly His Cys Pro Cys Pro Ser Asp Lys Pro Thr Ser

```
                180                 185                 190

Thr Ser Arg Asn Val Lys Arg Ala Cys Ser Asp Leu Glu Phe Arg Glu
        195                 200                 205

Val Ala Asn Arg Leu Arg Asp Trp Phe Lys Ala Leu His Glu Ser Gly
    210                 215                 220

Ser Gln Asn Lys Lys Thr Lys Thr Leu Leu Arg Pro Glu Arg Ser Arg
225                 230                 235                 240

Phe Asp Thr Ser Ile Leu Pro Ile Cys Lys Asp Ser Leu Gly Trp Met
                245                 250                 255

Phe Asn Arg Leu Asp Thr Asn Tyr Asp Leu Leu Leu Asp Gln Ser Glu
            260                 265                 270

Leu Arg Ser Ile Tyr Leu Asp Lys Asn Glu Gln Cys Thr Lys Ala Phe
        275                 280                 285

Phe Asn Ser Cys Asp Thr Tyr Lys Asp Ser Leu Ile Ser Asn Asn Glu
    290                 295                 300

Trp Cys Tyr Cys Phe Gln Arg Gln Gln Asp Pro Pro Cys Gln Thr Glu
305                 310                 315                 320

Leu Ser Asn Ile Gln Lys Arg Gln Gly Val Lys Lys Leu Leu Gly Gln
                325                 330                 335

Tyr Ile Pro Leu Cys Asp Glu Asp Gly Tyr Tyr Lys Pro Thr Gln Cys
            340                 345                 350

His Gly Ser Val Gly Gln Cys Trp Cys Val Asp Arg Tyr Gly Asn Glu
        355                 360                 365

Val Met Gly Ser Arg Ile Asn Gly Val Ala Asp Cys Ala Ile Asp Phe
    370                 375                 380

Glu Ile Ser Gly Asp Phe Ala Ser Gly Asp Phe His Glu Trp Thr Asp
385                 390                 395                 400

Asp Glu Asp Asp Glu Asp Asp Ile Met Asn Asp Glu Asp Glu Ile Glu
                405                 410                 415

Asp Asp Asp Glu Asp Glu Gly Asp Asp Asp Asp Gly Gly Asp Asp His
            420                 425                 430

Asp Val Tyr Ile
        435
```

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4130 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GGATTCGAAG TTTAAGAAAC TGCATTTTAA AGTGCCCAAA GTTTCATTTT CTTCTACCAA 60

AACTCCTAAA GATAGTTTAG TCCCAGGTGC AAAGTCTAGC ATAGGTCTTT CCACGATTCC 120

TTTATCATCT TCAGAATGCT CAAGTTTTGA ATTACAACAG GTTTCGGCTT GTTCAGAGCC 180

ATCCATGCAG ATGCCTAAGG TGGGTTTTGC TGGGTTTCCA TCATCCCGGC TTGATCTCAC 240

TGGTCCTCAC TTTGAATCTT CTATTCTCTC TCCCTGTGAG GATGTTACAC TTACAAAATA 300

CCAGGTGACT GTTCCCCAGA GCTGCCTTGG CCCCTGAGCT TGCTCTGGAA ATTCCTTCTG 360

GGTCTCAGGC TGATATTCCT CTTCCCAAGA CAGAGTGCTC CACTGAMCTG CAGCCTCCAG 420

ARGGAGTTCC AACATCTCAA GCTGAGAGTC ACTCTGGCCC ACTGAATTCC ATGATTCCTG 480

TTTCTCTTGG TCAGGTGTCT TTTCCTAAAT TCTATAAACC AAAGTTTGTG TTTTCAGTCC 540

CCCAAATGGC AGTTCCTGAG GGAGACCTAC ATGCAGCAGT GGGTGCCCCA GTCATGTYTC 600

YTCTTAGCCC TTGGAGAAAG AGTGCAGTGC CCCTTGCCAA GCACCCAGYT GCCATCCCCA 660

GGCACCTGTG TGTCCCAGGG CCCAGAAGAG CTTGTGGCCT CCTTGCAGAC ATCAGTAGTG 720

GCCCYTGGAG AAGCCCCTTC TGAAGATGCT GACCACGAAG GGAAAGGGAG TCCCTTGAAA 780

ATGCCTAAGA TTAAGCTTCC ATCATTTAGG TGGTCCCCGA AGAAGGAAAC AGGGCCAAAG 840

GTGGACCCAG AATGCAGCGT GGAGGACTCA AAACTCAGCC TGGTTTTAGA CAAGGATGAA 900

GTGGCCCCGC AGTCTGCCAT CCACATGGAT CTGCCTCCTG AGAGGGATGG AGAGAAGGGG 960

AGGAGCACAA AGCCTGGCTT TGCCATGCCA AAACTTGCAC TTCCCAAAAT GAAGGCTTCT 1020

AAGAGTGGGG TCAGCCTGCC ACAGAGAGAC GTGGATCCTT CCCTTTCTAG TGCCACAGCA 1080

GGGGGTAGCT TTCAAGACAC AGAAAAGGCC AGCAGTGACG GTGGTAGGGG AGGACTTGGT 1140

GCAACAGCAA GTGCCACAGG AAGTGAGGGT GTGAACCTCC ACCGGCCACA GGTCCACATT 1200

CCCAGTTTGG GCTTTGCCAA ACCTGATCTC AGATCCTCCA AGGCCAAGGT GGAGGTGAGC 1260

CAGCCTGAAG CTGACCTGCC TCTTCCCAAA CATGATCTGT CTACCGAAGG TGACAGCAGA 1320

GGATGTGGGC TCGAGGATGT CCCAGTGAGC CAGCCTTGTG GGGAGGGGAT AGCCCCCACA 1380

CCTGAAGATC CCCTCCAGCC ATCCTGTAGA AAACCAGATG CTGAAGTCCT CACAGTGGAA 1440

AGCCCAGAGG AGGAAGCCAT GACCAAGGAC TCGCAGGAAA GCTGGTTTAA AATGCCCAAG 1500

TTCCGCATGC CCAGCCTTAG GCGCTCTTTC AGGGACAGAG GCGGGGCTGG AAAGCTGGAA 1560

GTGGCTCAGA CACAGGCACC GGCAGCAACA GGGGGTGAAG CAGCAGCTAA AGTCAAAGAG 1620

TTCCTTGTTT CTGGGTCAAA CGTGGAGGCA GCTATGTCCC TACAGCTCCC AGAGGCAGAT 1680

GCAGAAGTGA CAGCTTCTGA GAGCAAATCA TCCACAGATA TTCTAAGGTG TGATCTTGAC 1740

AGCACAGGCT TGAAGCTGCA CCTTTCCACT GCTGGGATGA CTGGGGATGA GCTTTCCACT 1800

TCTGAGGTCA GGATCCATCC ATCCAAAGGA CCTCTCCCTT TTCAGATGCC TGGCATGAGG 1860

CTTCCAGAAA CCCAGGTTCT TCCAGGAGAA ATAGATGAGA CTCCTCTTTC CAAGCCAGGA 1920

CATGACCTTG CCAGCATGGA GGATAAAACA GAGAAATGGT CTTCCCAGCC TGAAGGTCCA 1980

CTTAAATTGA AAGCTTCAAG TACTGATATG CCATCCCAGA TTTCTGTGGT TAATGTGGAT 2040

CAACTGTGGG AAGATTCTGT CCTAACTGTC AAATTCCCCA AATTAATGGT ACCAAGGTTC 2100

TCCTTCGCTG CCCCCAGCTC AGAGGATGAT GTGTTCATCC CCACTGTGAG GGAAGTGCAG 2160

TGTCCAGAGG CCAATATTGA TACAGCCCTT TGTAAGGAAA GTCCGGGGCT CTGGGGAGCC 2220

AGCATCCTGA AGGCAGGTGC TGGGGTCCCT GGGGAGCAGC CTGTGGACCT TAACCTGCCT 2280

TTGGAAGCTC CCCCAATTTC AAAGGTCAGA GTGCATATTC AGGGTGCTCA GGTTGAAAGT 2340

CAAGAGGTCA CTATACACAG CATAGTGACA CCAGAGTTTG TAGATCTCTC AGTACCCAGG 2400

ACTTTTTCCA CTCAGATTGT GCGGGAATCA GAGATCCCCA CGTCAGAGAT TCAAACACCT 2460

TCGTACGGAT TTTCCTTATT AAAAGTGAAA ATCCCAGAGC CCCACACGCA GGCTAGAGTG 2520

TACACAACAA TGACTCAACA CTCTAGGACT CAGGAGGGCA CAGAAGAGGC TCCCATACAA 2580

GCCACCCCAG GAGTAGACTC CATTTCTGGA GATCTCCAGC CTGACACTGG AGAACCATTT 2640

GAGATGATCT CTTCCAGCGT CAATGTACTG GGACAGCAAA CACTCACATT TGAAGTTCCT 2700

TCTGGCCACC AGCTTGCAGA CAGCTGTTCA GATGAGGAGC CAGCAGAAAT TCTTGAGTTT 2760

CCCCCTGATG ATAGCCAAGA GGCAACCACA CCACTGGCAG ATGAAGGCAG GGCTCCAAAA 2820

GACAAACCAG AAAGTAAAAA ATCTGGTCTG CTCTGGTTTT GGCTTCCAAA CATTGGGTTT 2880

TCCTCTTCTG TTGATGAGAC AGGTGTTGAT TCCAAAAATG ACGTCCAGAG ATCTGCTCCC 2940

ATTCAAACAC AGCCTGAGGC ACGACCAGAG GCAGAACTGC CTAAAAAACA GGAGAAGGCA 3000

GGCTGGTTCC GATTTCCCAA ATTAGGGTTC TCCTCATCTC CTACCAAGAA AAGCAAAAGC 3060

ACCGAAGATG GGGCAGAGCT GGAAGAACAA AAACTTCAAG AAGAAACAAT CACGTTTTTC 3120

GATGCCCGAG AAAGTTTCTC CCCTGAAGAG AAGGAAGAGG GTGAACTGAT CGGGCCTGTG 3180

GGCACTGGGC TGGACTCCAG AGTGATGGTG ACATCCGCGG CAAGAACAGA GTTAATCCTG 3240

CCCGAGCAGG ACAGAAAAGC TGACGATGAA AGCAAAGGGT CAGGCCTGGG ACCAAATGAA 3300

GGCTGAGAGG TATGGCTCAT CGGTACAAGA GAGATGCAAA AAACTAAGTT GGAAAGTAAA 3360

GGCTACACAC ACATATGGAG CACCCCATCC CACAGCACAT TACATCCACC TCACTTCACA 3420

GAACGGAGAA CAGAGCAGAA ATGACCAGAA CACCTTTGTC ACCATCACAC AGCCCTCCTA 3480

AAATGGAACC AAAGCTTCCC AGCTCCCTCA AAGCTTTGGA TGCAAAGAAG GCACCCTGAC 3540

TTCCACAAGA CACCAGAATT CACACGGTAC TCAGAGGCAC TGCTGGGGAA GTTTGTTGGT 3600

CTTTATTAGA TAAATTTCCA GAGACCTGTC CATAATACCC AACAGAACAT GACTGTTTCT 3660

TTGAGGAAAG GGTTATAATG TCTGTGGTGT ACAAGTCGTT TTTGGTATAA CTTCTTTCCT 3720

GCTGCTGCTG CTTCCCGGCA AACATAGTTT TCCTATTTCA GGCAGAGTGC GGTATATTCC 3780

AGGAAACACT GTTTCCTACT CACTTAGCTT ACTTCTTTGT TGAATGCCTC ACTAATGGCA 3840

AGTTTCAAGA TGTTTTGGGT GACAATGCAC ACATGCTGGG CAAAAGGGTG ATGGCCAGTG 3900

GCTGGCAGCT GGGCCAGCAG AAGCTAGGAC ATCTGTGAGT TGTCATTCTC ATCTATCCAT 3960

GTCCACTGGC CTGCCAGCAT CCGCCAGTGC CTTGCCAGTG TGCACGGTCC CACACTGTGG 4020

CCCCTGAGTC CCCTAATGTA CACGCTGCAG CCAGAATGCA GATGGAGCTG GCTTGGCTGT 4080

TCCCTGGATG GGCAATAAAG AAAGTGCTGC ATCCCAAAAA AAAAAAAAAA 4130

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 911 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Met Gln Gln Trp Val Pro Gln Ser Cys Xaa Xaa Leu Ala Leu Gly Glu
1 5 10 15

Arg Val Gln Cys Pro Leu Pro Ser Thr Gln Leu Pro Ser Pro Gly Thr

```
            20                  25                  30

Cys Val Ser Gln Gly Pro Glu Glu Leu Val Ala Ser Leu Gln Thr Ser
        35                  40                  45

Val Val Ala Xaa Gly Glu Ala Pro Ser Glu Asp Ala Asp His Glu Gly
    50                  55                  60

Lys Gly Ser Pro Leu Lys Met Pro Lys Ile Lys Leu Pro Ser Phe Arg
65                  70                  75                  80

Trp Ser Pro Lys Lys Glu Thr Gly Pro Lys Val Asp Pro Glu Cys Ser
                85                  90                  95

Val Glu Asp Ser Lys Leu Ser Leu Val Leu Asp Lys Asp Glu Val Ala
            100                 105                 110

Pro Gln Ser Ala Ile His Met Asp Leu Pro Pro Glu Arg Asp Gly Glu
        115                 120                 125

Lys Gly Arg Ser Thr Lys Pro Gly Phe Ala Met Pro Lys Leu Ala Leu
    130                 135                 140

Pro Lys Met Lys Ala Ser Lys Ser Gly Val Ser Leu Pro Gln Arg Asp
145                 150                 155                 160

Val Asp Pro Ser Leu Ser Ser Ala Thr Ala Gly Gly Ser Phe Gln Asp
                165                 170                 175

Thr Glu Lys Ala Ser Ser Asp Gly Gly Arg Gly Gly Leu Gly Ala Thr
            180                 185                 190

Ala Ser Ala Thr Gly Ser Glu Gly Val Asn Leu His Arg Pro Gln Val
        195                 200                 205

His Ile Pro Ser Leu Gly Phe Ala Lys Pro Asp Leu Arg Ser Ser Lys
    210                 215                 220

Ala Lys Val Glu Val Ser Gln Pro Glu Ala Asp Leu Pro Leu Pro Lys
225                 230                 235                 240

His Asp Leu Ser Thr Glu Gly Asp Ser Arg Gly Cys Gly Leu Glu Asp
                245                 250                 255

Val Pro Val Ser Gln Pro Cys Gly Glu Gly Ile Ala Pro Thr Pro Glu
            260                 265                 270

Asp Pro Leu Gln Pro Ser Cys Arg Lys Pro Asp Ala Glu Val Leu Thr
        275                 280                 285

Val Glu Ser Pro Glu Glu Glu Ala Met Thr Lys Asp Ser Gln Glu Ser
    290                 295                 300

Trp Phe Lys Met Pro Lys Phe Arg Met Pro Ser Leu Arg Arg Ser Phe
305                 310                 315                 320
```

```
Arg Asp Arg Gly Gly Ala Gly Lys Leu Glu Val Ala Gln Thr Gln Ala
                325                 330                 335

Pro Ala Ala Thr Gly Gly Glu Ala Ala Ala Lys Val Lys Glu Phe Leu
            340                 345                 350

Val Ser Gly Ser Asn Val Glu Ala Ala Met Ser Leu Gln Leu Pro Glu
        355                 360                 365

Ala Asp Ala Glu Val Thr Ala Ser Glu Ser Lys Ser Ser Thr Asp Ile
    370                 375                 380

Leu Arg Cys Asp Leu Asp Ser Thr Gly Leu Lys Leu His Leu Ser Thr
385                 390                 395                 400

Ala Gly Met Thr Gly Asp Glu Leu Ser Thr Ser Glu Val Arg Ile His
                405                 410                 415

Pro Ser Lys Gly Pro Leu Pro Phe Gln Met Pro Gly Met Arg Leu Pro
            420                 425                 430

Glu Thr Gln Val Leu Pro Gly Glu Ile Asp Glu Thr Pro Leu Ser Lys
        435                 440                 445

Pro Gly His Asp Leu Ala Ser Met Glu Asp Lys Thr Glu Lys Trp Ser
    450                 455                 460

Ser Gln Pro Glu Gly Pro Leu Lys Leu Lys Ala Ser Ser Thr Asp Met
465                 470                 475                 480

Pro Ser Gln Ile Ser Val Val Asn Val Asp Gln Leu Trp Glu Asp Ser
                485                 490                 495

Val Leu Thr Val Lys Phe Pro Lys Leu Met Val Pro Arg Phe Ser Phe
            500                 505                 510

Ala Ala Pro Ser Ser Glu Asp Asp Val Phe Ile Pro Thr Val Arg Glu
        515                 520                 525

Val Gln Cys Pro Glu Ala Asn Ile Asp Thr Ala Leu Cys Lys Glu Ser
    530                 535                 540

Pro Gly Leu Trp Gly Ala Ser Ile Leu Lys Ala Gly Ala Gly Val Pro
545                 550                 555                 560

Gly Glu Gln Pro Val Asp Leu Asn Leu Pro Leu Glu Ala Pro Pro Ile
                565                 570                 575

Ser Lys Val Arg Val His Ile Gln Gly Ala Gln Val Glu Ser Gln Glu
            580                 585                 590

Val Thr Ile His Ser Ile Val Thr Pro Glu Phe Val Asp Leu Ser Val
        595                 600                 605

Pro Arg Thr Phe Ser Thr Gln Ile Val Arg Glu Ser Glu Ile Pro Thr
```

```
    610                 615                 620

Ser Glu Ile Gln Thr Pro Ser Tyr Gly Phe Ser Leu Leu Lys Val Lys
625                 630                 635                 640

Ile Pro Glu Pro His Thr Gln Ala Arg Val Tyr Thr Thr Met Thr Gln
                645                 650                 655

His Ser Arg Thr Gln Glu Gly Thr Glu Glu Ala Pro Ile Gln Ala Thr
            660                 665                 670

Pro Gly Val Asp Ser Ile Ser Gly Asp Leu Gln Pro Asp Thr Gly Glu
        675                 680                 685

Pro Phe Glu Met Ile Ser Ser Ser Val Asn Val Leu Gly Gln Gln Thr
    690                 695                 700

Leu Thr Phe Glu Val Pro Ser Gly His Gln Leu Ala Asp Ser Cys Ser
705                 710                 715                 720

Asp Glu Glu Pro Ala Glu Ile Leu Glu Phe Pro Pro Asp Asp Ser Gln
                725                 730                 735

Glu Ala Thr Thr Pro Leu Ala Asp Glu Gly Arg Ala Pro Lys Asp Lys
            740                 745                 750

Pro Glu Ser Lys Lys Ser Gly Leu Leu Trp Phe Trp Leu Pro Asn Ile
        755                 760                 765

Gly Phe Ser Ser Ser Val Asp Glu Thr Gly Val Asp Ser Lys Asn Asp
    770                 775                 780

Val Gln Arg Ser Ala Pro Ile Gln Thr Gln Pro Glu Ala Arg Pro Glu
785                 790                 795                 800

Ala Glu Leu Pro Lys Lys Gln Glu Lys Ala Gly Trp Phe Arg Phe Pro
                805                 810                 815

Lys Leu Gly Phe Ser Ser Ser Pro Thr Lys Lys Ser Lys Ser Thr Glu
            820                 825                 830

Asp Gly Ala Glu Leu Glu Glu Gln Lys Leu Gln Glu Glu Thr Ile Thr
        835                 840                 845

Phe Phe Asp Ala Arg Glu Ser Phe Ser Pro Glu Glu Lys Glu Glu Gly
    850                 855                 860

Glu Leu Ile Gly Pro Val Gly Thr Gly Leu Asp Ser Arg Val Met Val
865                 870                 875                 880

Thr Ser Ala Ala Arg Thr Glu Leu Ile Leu Pro Glu Gln Asp Arg Lys
                885                 890                 895

Ala Asp Asp Glu Ser Lys Gly Ser Gly Leu Gly Pro Asn Glu Gly
            900                 905                 910
```

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4142 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GCTCGTCTCG CCGGGCTGTT CGCGGGCAGG CCCTGCCCTG AAGGGACGAA TCGGCTTGGA 60

GCGCGGGAGG TGGAGTCGGC CCCGGCGGTC GCTCCCTGGA CCCAACCCGA GGCTGACCCA 120

KGCCCCTGCC CATGCGGGGC GCCCCTGGCT CGGAAGAGTC CCCCGGGCCG GGAGCAGCTC 180

CAGGCAGCGG CCCCGGAGGA AGAGGAAGAA GGGACAGTGC TCAGCTTGGG GGACCCGGAC 240

CCTCGCCGCG GCATTTGGAG CCGGGGGCAG TCCCGAACTC TGTGCTTGGC ACCGCCGCTC 300

CGAGTAGGGC AGCGCCTGCC GGGACTCTGA CCCGGACCCC CTGCGCCTCG TAGGCGGCGG 360

CGCCGCCGCG CCACCCTGTT CTTCCGTGTC TCCCTCTGCC TGGCGGCAGT CACGGCCAAG 420

AGAGTATTAT GAGGGAGGCC GAGGACTTCA TGCTCCGGAC AGAGAAACGG CGCTGGGATT 480

AGGGATTGCC ACTTCTGAGA GGATGCTGGG AATCTGCAGG GGGAGACGGA AATTCTTGGC 540

TGCCTCGTTG AGTCTTCTCT GCATCCCAGC CATCACCTGG ATTTACCTGT TTTCTGGGAG 600

CTTCGAAGAT GGAAAGCCCG TGTCTCTGTC ACCGCTGGAG TCCCAGGCAC ACAGCCCCAG 660

GTACACGGCC TCCAGCCAGC GGGAGCGCGA GAGCCTGGAG GTGCGCATGC GCGAGGTGGA 720

GGAGGAGAAC CGCGCCCTCC GCAGGCAGCT CAGCCTGGCC CAGGGCCGAG CCCCATCCCA 780

TCGCCGAGGC AACCACTCCA AGACCTACTC CATGGAGGAG GGCACTGGAG ACAGCGAGAA 840

CCTTCGGGCT GGCATCGTGG CAGGCAACAG CTCCGAGTGT GGGCAGCAGC CGGTCGTGGA 900

GAAATGCGAG ACAATCCACG TTGCTATTGT CTGCGCCGGA TACAATGCCA GCCGGGATGT 960

CGTCACCCTG GTCAAATCCG TCCTGTTCCA TAGACGGAAC CCTCTGCACT TCCACCTTAT 1020

TGCTGACTCC ATTGCGGAGC AGATCCTGGC CACGCTCTTC CAGACCTGGA TGGTGCCCGC 1080

TGTGCGTGTG GACTTCTACA ATGCAGACGA GCTCAAGTCT GAAGTTTCCT GGATCCCCAA 1140

TAAACATTAC TCTGGGATTT ATGGTCTGAT GAAGCTTGTC CTGACCAAGA CTCTTCCTGC 1200

CAACCTGGAG AGAGTCATCG TCCTTGACAC GGATATCACC TTTGCCACTG ACATTGCAGA 1260

GCTGTGGGCT GTGTTCCACA AGTTCAAAGG TCAGCAAGTC CTGGGCTTGG TGGAGAACCA 1320

GAGTGACTGG TACCTTGGAA ACCTGTGGAA AAATCACCGC CCATGGCCAG CCCTTGGAAG 1380

AGGCTACAAC ACAGGGGTGA TCCTGTTACT TCTGGATAAG CTGCGGAAGA TGAAATGGGA 1440

GCAGATGTGG AGGCTGACCG CAGAGAGGGA GCTCATGGGC ATGCTCTCTA CATCCTTAGC 1500

TGACCAGGAT ATTTTCAATG CCGTCATCAA ACAAAACCCC TTCCTTGTGT ACCAGCTCCC 1560

CTGCTTCTGG AATGTGCAGC TGTCAGACCA CACCCGCTCC GAGCAGTGCT ACAGAGACGT 1620

GTCTGATCTA AAGGTCATTC ACTGGAACTC CCCCAAGAAG CTCCGGGTGA AGAACAAGCA 1680

TGTGGAGTTT TTTCGCAACC TCTACCTGAC CTTCCTGGAG TATGACGGCA ATCTTCTGAG 1740

GCGGGAACTG TTTGGCTGCC CCAGTGAGGC TGATGTCAAC AGTGAAAACC TCCAGAAGCA 1800

GCTGTCTGAG CTGGACGAGG ACGACCTGTG CTATGAGTTC CGGCGAGAGC GCTTCACTGT 1860

CCACCGCACC CACCTGTACT TCCTGCACTA CGAGTATGAG CCTGCAGCAG ACAGCACGGA 1920

CGTCACCCTG GTCGCTCAGC TGTCCATGGA CAGGCTCCAG ATGCTGGAGG CCATCTGCAA 1980

GCACTGGGAG GGGCCCATCA GCCTGGCCCT CTACCTGTCA GACGCCGAGG CCCAGCAGTT 2040

CCTCCGCTAC GCACAGGGCT CTGAGGTGCT TATGAGCCGC CACAACGTGG GCTACCACAT 2100

CGTGTACAAG GAGGGCCAGT TCTACCCCGT GAACCTGCTG CGCAACGTGG CCATGAAGCA 2160

CATCAGCACT CCCTACATGT TCCTGTCTGA CATTGACTTC CTGCCCATGT ATGGGCTCTA 2220

TGAGTACCTC AGGAAGTCTG TCATCCAGCT CGATCTTGCC AACACCAAGA AAGCAATGAT 2280

TGTCCCCGCG TTCGAGACAC TGCGCTACCG GCTGTCCTTC CCCAAGTCAA AAGCGGAGTT 2340

GCTGTCAATG CTGGACATGG GGACCCTCTT CACATTCAGG TACCACGTCT GGACGAAAGG 2400

CCACGCACCC ACAAACTTCG CCAAGTGGCG GACCGCCACC ACGCCTTACC GGGTTGAGTG 2460

GGAGGCCGAT TTTGAGCCGT ATGTTGTTGT GAGACGTGAC TGCCCGGAGT ACGACCGGAG 2520

GTTTGTAGGC TTTGGCTGGA ACAAAGTGGC TCATATCATG GAGCTGGATG TGCAGGAGTA 2580

TGAGTTCATT GTGCTGCCCA ACGCCTACAT GATCCACATG CCTCATGCCC CCAGCTTCGA 2640

CATTACCAAG TTCCGTTCCA ACAAGCAATA CCGCATCTGT CTCAAAACCC TCAAGGAAGA 2700

GTTTCAGCAG GACATGTCCC GCCGCTACGG CTTTGCTGCC CTGAAATATC TCACAGCCGA 2760

GAACAACAGC TAGCACCAAG AAGCCCACCA CTAGGGGGAG ACATGCTGTA GGGGAAGTGC 2820

CACTCGCTGT TTGGGGCCCG GCCTTCAAAT TCAAAATTGA GCCATGCTTT TTCGGTTTGT 2880

TTTTATTTAT CTCTTTGGCC CAGCCAAGCT GCCCTCACTA CAGAGACCTT GGACAAGGAT 2940

CCAGCCAGTC CCTCTCTGCC CCACAACCCT GCATTCCCAG AGGTTAGCTA TGCAGCCCAC 3000

CTAGATGAGT CTCTTCAAGA ATGGGAAATC AAGGGGTGAC AGGGAGTAAA AGGGTTATCA 3060

TCTTACTGCA AAGCCACAAG ATCAGGGCAG GGCTTTAGGA TGTTCTGGAT GCTTTTTAAT 3120

AATTATGCTT CCCATCATAA CTGGGGAGAA AGGGAAGTCA GGGTTCTAGG GGTTATTCGT 3180

CCCAGGAAAT AGAAGTGAAA TTGTCTTTAT TAAGTGAAAA CTTTCCCCTT TGCCCTGCAA 3240

TGTAGCTGGG CATTCAAACG GAGGGCAAAC CGATGATCTA AACCAACCAC TTGGAAAAAC 3300

CCAATGGGGA CATTGTAACC AGAGGGTCCT GGAGGTGGGG TTGATGGGTT TCCTTATCCC 3360

CAAAGTCACT CCTGTTTTGT TTTGTTTTTC TTTGGGGGTT TTGTTTATTT TTGGGGCTGG 3420

CAATCCAAAA TAGAAAATCT GATCCTTTGA GGCTCTAAAG GAAAATCAGC TGCCTCTACC 3480

AACCACCCTC TATCAGCAGT GGCCCAGGAA GGAGGTCAAG CATCTTCGGC CGATATTTAA 3540

ACATGGGCAG CTTCCTTCAG GATGATCACC GAGGCTCCCG TGACTTTGAA CTCCCTACTC 3600

TCCAGAATCC AGGGGCTATA GCGATGGGGA CTGCGGAATT ACGAGGGCTG GCTGTTTTAC 3660

ACCGGTCACA TTTTCTATTG GCAGTGACTG ATTCATGGGA AAGGGCTTTG AAGGAACTAC 3720

TTCAGTGCAC ACACAAGGTA CGAACCTYTC AGGCCTTTCG AAGAACTTTC ATAATTCATG 3780

AAAGCCCAGT TYTGAAGATT CACGTATCCA TYTGGAGACC TACAGGAAGA AAGTGATTGG 3840

GTTCCTCTGG TTCTTGCCTG CTTCACTGTG GATGGGAAGA GGTGACAACC TCAGTCTCCC 3900

TTTGGGACCT GTCCAAGGGT AGGCAACCAC CTTCACCTTC ACACAGATTG AGGAGACACT 3960

GGACTTTTTA CCCATTTTCT TTAATYTTCA ATATTAATAT TGTGTTTACA TTGATGAGAA 4020

CAAGAGTTAA TGCCCTACCC TCTGCTGGGC TGTTTGTATT GAGTTGCAAT GTGACCAGCG 4080

AAAGCTGCAT TTAATAAATG AAAGTACAGA CTGAAAAAAA AAAAAAAAAA AAAAAAAAAA 4140

AA 4142

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 756 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

```
Met Leu Gly Ile Cys Arg Gly Arg Arg Lys Phe Leu Ala Ala Ser Leu
1               5                  10                  15

Ser Leu Leu Cys Ile Pro Ala Ile Thr Trp Ile Tyr Leu Phe Ser Gly
            20                  25                  30

Ser Phe Glu Asp Gly Lys Pro Val Ser Leu Ser Pro Leu Glu Ser Gln
        35                  40                  45

Ala His Ser Pro Arg Tyr Thr Ala Ser Ser Gln Arg Glu Arg Glu Ser
    50                  55                  60

Leu Glu Val Arg Met Arg Glu Val Glu Glu Glu Asn Arg Ala Leu Arg
65                  70                  75                  80

Arg Gln Leu Ser Leu Ala Gln Gly Arg Ala Pro Ser His Arg Arg Gly
                85                  90                  95

Asn His Ser Lys Thr Tyr Ser Met Glu Glu Gly Thr Gly Asp Ser Glu
            100                 105                 110

Asn Leu Arg Ala Gly Ile Val Ala Gly Asn Ser Ser Glu Cys Gly Gln
        115                 120                 125

Gln Pro Val Val Glu Lys Cys Glu Thr Ile His Val Ala Ile Val Cys
    130                 135                 140

Ala Gly Tyr Asn Ala Ser Arg Asp Val Val Thr Leu Val Lys Ser Val
145                 150                 155                 160

Leu Phe His Arg Arg Asn Pro Leu His Phe His Leu Ile Ala Asp Ser
                165                 170                 175

Ile Ala Glu Gln Ile Leu Ala Thr Leu Phe Gln Thr Trp Met Val Pro
            180                 185                 190

Ala Val Arg Val Asp Phe Tyr Asn Ala Asp Glu Leu Lys Ser Glu Val
        195                 200                 205

Ser Trp Ile Pro Asn Lys His Tyr Ser Gly Ile Tyr Gly Leu Met Lys
    210                 215                 220

Leu Val Leu Thr Lys Thr Leu Pro Ala Asn Leu Glu Arg Val Ile Val
225                 230                 235                 240

Leu Asp Thr Asp Ile Thr Phe Ala Thr Asp Ile Ala Glu Leu Trp Ala
                245                 250                 255

Val Phe His Lys Phe Lys Gly Gln Gln Val Leu Gly Leu Val Glu Asn
            260                 265                 270

Gln Ser Asp Trp Tyr Leu Gly Asn Leu Trp Lys Asn His Arg Pro Trp
```

275 280 285

Pro Ala Leu Gly Arg Gly Tyr Asn Thr Gly Val Ile Leu Leu Leu Leu
290 295 300

Asp Lys Leu Arg Lys Met Lys Trp Glu Gln Met Trp Arg Leu Thr Ala
305 310 315 320

Glu Arg Glu Leu Met Gly Met Leu Ser Thr Ser Leu Ala Asp Gln Asp
325 330 335

Ile Phe Asn Ala Val Ile Lys Gln Asn Pro Phe Leu Val Tyr Gln Leu
340 345 350

Pro Cys Phe Trp Asn Val Gln Leu Ser Asp His Thr Arg Ser Glu Gln
355 360 365

Cys Tyr Arg Asp Val Ser Asp Leu Lys Val Ile His Trp Asn Ser Pro
370 375 380

Lys Lys Leu Arg Val Lys Asn Lys His Val Glu Phe Phe Arg Asn Leu
385 390 395 400

Tyr Leu Thr Phe Leu Glu Tyr Asp Gly Asn Leu Leu Arg Arg Glu Leu
405 410 415

Phe Gly Cys Pro Ser Glu Ala Asp Val Asn Ser Glu Asn Leu Gln Lys
420 425 430

Gln Leu Ser Glu Leu Asp Glu Asp Asp Leu Cys Tyr Glu Phe Arg Arg
435 440 445

Glu Arg Phe Thr Val His Arg Thr His Leu Tyr Phe Leu His Tyr Glu
450 455 460

Tyr Glu Pro Ala Ala Asp Ser Thr Asp Val Thr Leu Val Ala Gln Leu
465 470 475 480

Ser Met Asp Arg Leu Gln Met Leu Glu Ala Ile Cys Lys His Trp Glu
485 490 495

Gly Pro Ile Ser Leu Ala Leu Tyr Leu Ser Asp Ala Glu Ala Gln Gln
500 505 510

Phe Leu Arg Tyr Ala Gln Gly Ser Glu Val Leu Met Ser Arg His Asn
515 520 525

Val Gly Tyr His Ile Val Tyr Lys Glu Gly Gln Phe Tyr Pro Val Asn
530 535 540

Leu Leu Arg Asn Val Ala Met Lys His Ile Ser Thr Pro Tyr Met Phe
545 550 555 560

Leu Ser Asp Ile Asp Phe Leu Pro Met Tyr Gly Leu Tyr Glu Tyr Leu
565 570 575

```
Arg Lys Ser Val Ile Gln Leu Asp Leu Ala Asn Thr Lys Lys Ala Met
            580                 585                 590

Ile Val Pro Ala Phe Glu Thr Leu Arg Tyr Arg Leu Ser Phe Pro Lys
        595                 600                 605

Ser Lys Ala Glu Leu Leu Ser Met Leu Asp Met Gly Thr Leu Phe Thr
    610                 615                 620

Phe Arg Tyr His Val Trp Thr Lys Gly His Ala Pro Thr Asn Phe Ala
625                 630                 635                 640

Lys Trp Arg Thr Ala Thr Thr Pro Tyr Arg Val Glu Trp Glu Ala Asp
                645                 650                 655

Phe Glu Pro Tyr Val Val Val Arg Arg Asp Cys Pro Glu Tyr Asp Arg
            660                 665                 670

Arg Phe Val Gly Phe Gly Trp Asn Lys Val Ala His Ile Met Glu Leu
        675                 680                 685

Asp Val Gln Glu Tyr Glu Phe Ile Val Leu Pro Asn Ala Tyr Met Ile
    690                 695                 700

His Met Pro His Ala Pro Ser Phe Asp Ile Thr Lys Phe Arg Ser Asn
705                 710                 715                 720

Lys Gln Tyr Arg Ile Cys Leu Lys Thr Leu Lys Glu Glu Phe Gln Gln
                725                 730                 735

Asp Met Ser Arg Arg Tyr Gly Phe Ala Ala Leu Lys Tyr Leu Thr Ala
            740                 745                 750

Glu Asn Asn Ser
        755
```

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1435 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

```
TGCACCGGTG GTCGGCTGTT GGGTGTGGAG TTTCCCAGCG CCCCTCGGGT CCGACCCTTT      60

GAGCGTTCTG CTCCGGCGCC AGCCTACCTC GCTCCTCGGC GCCATGACCA CAACCACCAC     120
```

CTTCAAGGGA GTCGACCCCA ACAGCAGGAA TAGCTCCCGA GTTTTGCGGC CTCCAGGTGG 180

TGGATCCAAT TTTTCATTAG GTTTTGATGA ACCAACAGAA CAACCTGTGA GGAAGAACAA 240

AATGGCCTCT AATATCTTTG GGACACCTGA AGAAAATCAA GCTTCTTGGG CCAAGTCAGC 300

AGGTGCCAAG TCTAGTGGTG GCAGGGAAGA CTTGGAGTCA TCTGGACTGC AGAGAAGGAA 360

CTCCTCTGAA GCAAGCTCCG GAGACTTCTT AGATCTGAAG GGAGAAGGTG ATATTCATGA 420

AAATGTGGAC ACAGACTTGC CAGGCAGCCT GGGGCAGAGT GAAGAGAAGC CCGTGCCTGC 480

TGCGCCTGTG CCCAGCCCGG TGGCCCCGGC CCCAGTGCCA TCCAGAAGAA ATCCCCCTGG 540

CGGCAAGTCC AGCCTCGTCT TGGGTTAGCT CTGACTGTCC TGAACGCTGT CGTTCTGTCT 600

GTTTCCTCCA TGCTTGTGAA CTGCACAACT TGAGCCTGAC TGTACATCTC TTGGATTTGT 660

TTCATTAAAA AGAAGCACTT TATGTACTGC TGTCTTTTTT TTTTTTTCTT TTGAAGAACA 720

GGTTTCTCTC TGTCCTTGAC TCTTGGGTCT GTGGGCCATG GCATGAGTGT TTTCTAGTAG 780

TAGATTGGAG GGAAAGCTTT GTGACACTTA GTACTGTGTT TTTAAGAAGA AATAATTTGG 840

TTCCAGATGT GTTAGAGGAT CTTTTGTACT GAGGTTTTTA ACACTTTACT TGGGTTTACC 900

AAGCCTCAAC TGGACAGACC ATAAACAGTC CACAGGCACC GTTCCTGCCA GGCCCCAACC 960

CACAGGGAGT CTCTCCGCAG AGCCTTCTTG GTGTTGCCCT AACTTGCCAG TGGCCTTTGC 1020

TCAGAGCCTC CTCCTGTGAC ATGTGAACAA TGAAGAGGCC TGCGCYTCCT GCCTTGCCGC 1080

CTGCAAAGCA AAGAAACTGC CTTTTATTTT TTAACCTTAA AAAGTAGCCA GATAGTAACA 1140

AGACTGGCTG GCTGATGAGC AAAGCYTTTG CTCTCACGCA GAGGAAGGCT TGGATGTACA 1200

ATGAAACTGC CTGGAACTAA AAGCAGTGAA GCAAGGGAGG CAATCACACT GAAGCGGGTC 1260

TTCCTCCAGG AACGGGGTCC CACAGGCGTG TTGTTTTAAA TAACCTGATG CTGTGTGCAT 1320

GATGCTGGTG CTTGACCATG AAAGGAAAGT CTCATCCTTA AAATGTGTTG TACTTCACAA 1380

TCCTGGACTG TTGCTTCAAG TAAACAATAT CCACATTTTG AAAAAAAAAA AAAAA 1435

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 154 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Met Thr Thr Thr Thr Thr Phe Lys Gly Val Asp Pro Asn Ser Arg Asn
1 5 10 15

Ser Ser Arg Val Leu Arg Pro Pro Gly Gly Gly Ser Asn Phe Ser Leu
20 25 30

Gly Phe Asp Glu Pro Thr Glu Gln Pro Val Arg Lys Asn Lys Met Ala
35 40 45

Ser Asn Ile Phe Gly Thr Pro Glu Glu Asn Gln Ala Ser Trp Ala Lys
50 55 60

Ser Ala Gly Ala Lys Ser Ser Gly Gly Arg Glu Asp Leu Glu Ser Ser
65 70 75 80

Gly Leu Gln Arg Arg Asn Ser Ser Glu Ala Ser Ser Gly Asp Phe Leu
85 90 95

Asp Leu Lys Gly Glu Gly Asp Ile His Glu Asn Val Asp Thr Asp Leu
100 105 110

Pro Gly Ser Leu Gly Gln Ser Glu Glu Lys Pro Val Pro Ala Ala Pro
115 120 125

Val Pro Ser Pro Val Ala Pro Ala Pro Val Pro Ser Arg Arg Asn Pro
130 135 140

Pro Gly Gly Lys Ser Ser Leu Val Leu Gly
145 150

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1904 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

CAGCGTCGCG CGCGCTACCA CACCCAGGTT CGGCCCGTAG GCGTCTGGCA GCCCGGCGCC 60

ATCTTCATCG AGCGCCATGG CCGCAGCCTG CGGGCCGGGA GCGGCCGGGT ACTGCTTGCT 120

CCTCGGCTTG CATTTGTTTC TGCTGACCGC GGGCCCTGCC CTGGGCTGGA ACGACCCTGA 180

CAGAATGTTG CTGCGGGATG TAAAAGCTCT TACCCTCCAC TATGACCGCT ATACCACCTC 240

CCGCAGGCTG GATCCCATCC CACAGTTGAA ATGTGTTGGA GGCACAGCTG GTTGTGATTC 300

TTATACCCCA AAAGTCATAC AGTGTCAGAA CAAAGGCTGG GATGGGTATG ATGTACAGTG 360

GGAATGTAAG ACGGACTTAG ATATTGCATA CAAATTTGGA AAAACTGTGG TGAGCTGTGA 420

AGGCTATGAG TCCTCTGAAG ACCAGTATGT ACTAAGAGGT TCTTGTGGCT TGGAGTATAA 480

TTTAGATTAT ACAGAACTTG GCCTGCAGAA ACTGAAGGAG TCTGGAAAGC AGCACGGCTT 540

TGCCTCTTTC TCTGATTATT ATTATAAGTG GTCCTCGGCG GATTCCTGTA ACATGAGTGG 600

ATTGATTACC ATCGTGGTAC TCCTTGGGAT CGCCTTTGTA GTCTATAAGC TGTTCCTGAG 660

TGACGGGCAG TATTCTCCTC CACCGTACTC TGAGTATCCT CCATTTTCCC ACCGTTACCA 720

GAGATTCACC AACTCAGCAG GACCTCCTCC CCCAGGCTTT AAGTCTGAGT TCACAGGACC 780

ACAGAATACT GGCCATGGTG CAACTTCTGG TTTTGGCAGT GCTTTTACAG GACAACAAGG 840

ATATGAAAAT TCAGGACCAG GGTTCTGGAC AGGCTTGGGA ACTGGTGGAA TACTAGGATA 900

TTTGTTTGGC AGCAATAGAG CGGCAACACC CTTCTCAGAC TCGTGGTACT ACCCGTCCTA 960

TCCTCCCTCC TACCCTGGCA CGTGGAATAG GGCTTACTCA CCCCTTCATG GAGGCTCGGG 1020

CAGCTATTCG GTATGTTCAA ACTCAGACAC GAAAACCAGA ACTGCATCAG GATATGGTGG 1080

TACCAGGAGA CGATAAAGTA GAAAGTTGGA GTCAAACACT GGATGCAGAA ATTTTGGATT 1140

TTTCATCACT TTCTCTTTAG AAAAAAAGTA CTACCTGTTA ACAATTGGGA AAAGGGGATA 1200

TTCAAAAGTT CTGTGGTGTT ATGTCCAGTG TAGCTTTTTG TATTCTATTA TTTGAGGCTA 1260

AAAGTTGATG TGTGACAAAA TACTTATGTG TTGTATGTCA GTGTAACATG CAGATGTATA 1320

TTGCAGTTTT KGAAAGTGAT CATTACTGTG GAATGCTAAA AATACATTAA TTTCTAAAAC 1380

CTGTGATGCC CTAAGAAGCA TTAAGAATGA AGGTGTTGTA CTAATAGAAA CTAAGTACAG 1440

AAAATTTCAG TTTTAGGTGG TTGTAGCTGA TGAGTTATTA CCTCATAGAG ACTATAATAT 1500

TCTATTTGGT ATTATATTAT TTGATGTTTG CTGTTCTTCA AACATTTAAA TCAAGCTTTG 1560

GACTAATTAT GCTAATTTGT GAGTTCTGAT CACTTTTGAG CTCTGAAGCT TTGAATCATT 1620

CAGTGGTGGA GATGGCCTTC TGGTAACTGA ATATTACCTT CTGTAGGAAA AGGTGGAAAA 1680

TAAGCATCTA GAAGGTTGTT GTGAATGACT CTGTGCTGGC AAAAATGCTT GAAACCTCTA 1740

TATTTCTTTC GTTCATAAGA GGTAAAGGTC AAATTTTTCA ACAAAAGTCT TTTAATAACA 1800

AAAGCATGCA GTTCTCTGTG AAATCTCAAA TATTGTTGTA ATAGTCTGTT TCAATCTTAA 1860

AAAGAATCAA TAAAAACAAA CAAGGAAAAA AAAAAAAAAA AAAA 1904

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 339 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

```
Met Ala Ala Ala Cys Gly Pro Gly Ala Ala Gly Tyr Cys Leu Leu Leu
1               5                  10                  15

Gly Leu His Leu Phe Leu Leu Thr Ala Gly Pro Ala Leu Gly Trp Asn
            20                  25                  30

Asp Pro Asp Arg Met Leu Leu Arg Asp Val Lys Ala Leu Thr Leu His
        35                  40                  45

Tyr Asp Arg Tyr Thr Thr Ser Arg Arg Leu Asp Pro Ile Pro Gln Leu
    50                  55                  60

Lys Cys Val Gly Gly Thr Ala Gly Cys Asp Ser Tyr Thr Pro Lys Val
65                  70                  75                  80

Ile Gln Cys Gln Asn Lys Gly Trp Asp Gly Tyr Asp Val Gln Trp Glu
                85                  90                  95

Cys Lys Thr Asp Leu Asp Ile Ala Tyr Lys Phe Gly Lys Thr Val Val
            100                 105                 110

Ser Cys Glu Gly Tyr Glu Ser Ser Glu Asp Gln Tyr Val Leu Arg Gly
        115                 120                 125

Ser Cys Gly Leu Glu Tyr Asn Leu Asp Tyr Thr Glu Leu Gly Leu Gln
    130                 135                 140

Lys Leu Lys Glu Ser Gly Lys Gln His Gly Phe Ala Ser Phe Ser Asp
145                 150                 155                 160

Tyr Tyr Tyr Lys Trp Ser Ser Ala Asp Ser Cys Asn Met Ser Gly Leu
                165                 170                 175

Ile Thr Ile Val Val Leu Leu Gly Ile Ala Phe Val Val Tyr Lys Leu
            180                 185                 190

Phe Leu Ser Asp Gly Gln Tyr Ser Pro Pro Pro Tyr Ser Glu Tyr Pro
        195                 200                 205

Pro Phe Ser His Arg Tyr Gln Arg Phe Thr Asn Ser Ala Gly Pro Pro
```

210 215 220

Pro Pro Gly Phe Lys Ser Glu Phe Thr Gly Pro Gln Asn Thr Gly His
225 230 235 240

Gly Ala Thr Ser Gly Phe Gly Ser Ala Phe Thr Gly Gln Gln Gly Tyr
245 250 255

Glu Asn Ser Gly Pro Gly Phe Trp Thr Gly Leu Gly Thr Gly Gly Ile
260 265 270

Leu Gly Tyr Leu Phe Gly Ser Asn Arg Ala Ala Thr Pro Phe Ser Asp
275 280 285

Ser Trp Tyr Tyr Pro Ser Tyr Pro Pro Ser Tyr Pro Gly Thr Trp Asn
290 295 300

Arg Ala Tyr Ser Pro Leu His Gly Gly Ser Gly Ser Tyr Ser Val Cys
305 310 315 320

Ser Asn Ser Asp Thr Lys Thr Arg Thr Ala Ser Gly Tyr Gly Gly Thr
325 330 335

Arg Arg Arg

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1260 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

CTGTCTGGCG GCGGCAGCAT GGCGGCGGGG GCGGCTGAGG CAGCTGTAGC GGCCGTGGAG 60

GAGGTCGGCT CAGCCGGGCA GTTTGAGGAG CTGCTGCGCC TCAAAGCCAA GTCCCTCCTT 120

GTGGTCCATT TCTGGGCACC ATGGGCTCCA CAGTGTGCAC AGATGAACGA AGTTATGGCA 180

GAGTTAGCTA AAGAACTCCC TCAAGTTTCA TTTGTGAAGT TGGAAGCTGA AGGTGTTCCT 240

GAAGTATCTG AAAAATATGA AATTAGCTCT GTTCCCACTT TTCTGTTTTT CAAGAATTCT 300

CAGAAAATCG ACCGATTAGA TGGTGCACAT GCCCCAGAGT TGACCAAAAA AGTTCAGCGA 360

CATGCATCTA GTGGCTCCTT CCTACCCAGC GCTAATGAAC ATCTTAAAGA AGACCTCAGC 420

CTTCGCCTGA AAAAGCTGAC TCACGCTGCC CCCTGCATGC TGTTCATGAA GGGAACACCT 480

CAAGAACCAC GCTGTGGTTT CAGCAAGCAG ATGGTGGAAA TCCTTCACAA ACACAATATT 540

CAGTTCAGCA GCTTTGATAT CTTCTCAGAT GAAGAAGTTC GACAGGGGCT CAAAACGTAC 600

TCTAATTGGC CCACCTATCC TCAGCTCTAT GTTTCTGGAG AGCTAATAGG AGGACTTGAC 660

ATAATTAAGG AGCTGGAAGC ATCAGAAGAG CTGGACACGA TCTGTCCCAA AGCTCCCAAA 720

TTAGAGGAAA GGCTCAAAGT GCTGACAAAT AAAGCTTCTG TGATGCTCTT TATGAAAGGA 780

AACAAACAGG AAGCAAAATG TGGATTCAGC AAACAAATTC TGGAAATACT AAATAGTACT 840

GGTGTTGAAT ATGAAACATT CGATATATTG GAGGATGAAG AAGTTCGGCA AGGATTAAAA 900

GCTTACTCAA ATTGGCCAAC ATACCCTCAG CTGTATGTGA AAGGGGAGCT GGTGGGAGGA 960

TTGGATATTG TGAAGGAACT GAAAGAAAAT GGTGAATTGC TGCCTATACT GAGAGGAGAA 1020

AATTAATAAA TCTTAAACTT GGTGCCCAAC TATTGTAAGA AATATTTAAT TACATTGGGA 1080

GCAGTTCATG ATTTAGTCCT CAGAAATGGA CTAGGAATAG AAAATTCCTG CTTTCTCAGT 1140

TACATGTTTT GTGTATTTCA CAATGTCGTG CTAAATAAAT GTATGTTACA TTTTTTTCCC 1200

ACCAAAAATA GAATGCAATA AACATCTTCA AATTATTAAC AATAAAAAAA AAAAAAAAAA 1260

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 335 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Met Ala Ala Gly Ala Ala Glu Ala Ala Val Ala Ala Val Glu Glu Val
1 5 10 15

Gly Ser Ala Gly Gln Phe Glu Glu Leu Leu Arg Leu Lys Ala Lys Ser
20 25 30

Leu Leu Val Val His Phe Trp Ala Pro Trp Ala Pro Gln Cys Ala Gln
35 40 45

Met Asn Glu Val Met Ala Glu Leu Ala Lys Glu Leu Pro Gln Val Ser
50 55 60

Phe Val Lys Leu Glu Ala Glu Gly Val Pro Glu Val Ser Glu Lys Tyr
65 70 75 80

Glu Ile Ser Ser Val Pro Thr Phe Leu Phe Phe Lys Asn Ser Gln Lys
85 90 95

Ile Asp Arg Leu Asp Gly Ala His Ala Pro Glu Leu Thr Lys Lys Val
100 105 110

Gln Arg His Ala Ser Ser Gly Ser Phe Leu Pro Ser Ala Asn Glu His
115 120 125

Leu Lys Glu Asp Leu Ser Leu Arg Leu Lys Lys Leu Thr His Ala Ala
130 135 140

Pro Cys Met Leu Phe Met Lys Gly Thr Pro Gln Glu Pro Arg Cys Gly
145 150 155 160

Phe Ser Lys Gln Met Val Glu Ile Leu His Lys His Asn Ile Gln Phe
165 170 175

Ser Ser Phe Asp Ile Phe Ser Asp Glu Glu Val Arg Gln Gly Leu Lys
180 185 190

Thr Tyr Ser Asn Trp Pro Thr Tyr Pro Gln Leu Tyr Val Ser Gly Glu
195 200 205

Leu Ile Gly Gly Leu Asp Ile Ile Lys Glu Leu Glu Ala Ser Glu Glu
210 215 220

Leu Asp Thr Ile Cys Pro Lys Ala Pro Lys Leu Glu Glu Arg Leu Lys
225 230 235 240

Val Leu Thr Asn Lys Ala Ser Val Met Leu Phe Met Lys Gly Asn Lys
245 250 255

Gln Glu Ala Lys Cys Gly Phe Ser Lys Gln Ile Leu Glu Ile Leu Asn
260 265 270

Ser Thr Gly Val Glu Tyr Glu Thr Phe Asp Ile Leu Glu Asp Glu Glu
275 280 285

Val Arg Gln Gly Leu Lys Ala Tyr Ser Asn Trp Pro Thr Tyr Pro Gln
290 295 300

Leu Tyr Val Lys Gly Glu Leu Val Gly Gly Leu Asp Ile Val Lys Glu
305 310 315 320

Leu Lys Glu Asn Gly Glu Leu Leu Pro Ile Leu Arg Gly Glu Asn
325 330 335

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1152 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

ACTTTTTGCG ATGCCTACTG GAGACTTTGA TTCGAAGCCC AGTTGGGCCG ACCAGGTGGA 60

GGAGGAGGGG GAGGACGACA AATGTGTCAC CAGCGAGCTC CTCAAGGGGA TCCCTCTGGC 120

CACAGGTGAC ACCAGCCCAG AGCCAGAGCT ACTGCCGGGA GCTCCACTGC CGCCTCCCAA 180

GGAGGTCATC AACGGAAACA TAAAGACAGT GACAGAGTAC AAGATAGATG AGGATGGCAA 240

GAAGTTCAAG ATTGTCCGCA CCTTCAGGAT TGAGACCCGG AAGGCTTCAA AGGCTGTCGC 300

AAGGAGGAAG AACTGGAAGA AGTTCGGGAA CTCAGAGTTT GACCCCCCCG GACCCAATGT 360

GGCCACCACC ACTGTCAGTG ACGATGTCTC TATGACGTTC ATCACCAGCA AAGAGGACCT 420

GAACTGCCAG GAGGAGGAGG ACCCTATGAA CAAACTCAAG GGCCAGAAGA TCGTGTCCTG 480

CCGCATCTGC AAGGGCGACC ACTGGACCAC CCGCTGCCCC TACAAGGATA CGCTGGGGCC 540

CATGCAGAAG GAGCTGGCCG AGCAGCTGGG CCTGTCTACT GGCGAGAAGG AGAAGCTGCC 600

GGGAGAGCTA GAGCCGGTGC AGGCCACGCA GAACAAGACA GGGAAGTATG TGCCGCCGAG 660

CCTGCGCGAC GGGGCCAGCC GCCGCGGGGA GTCCATGCAG CCCACCCGCA GAGCCGACGA 720

CAACGCCACC ATCCGTGTCA CCAACTTGTC AGAGGACACG CGTGAGACCG ACCTGCAGGA 780

GCTCTTCCGG CCTTTCGGCT CCATCTCCCG CATCTACCTG GCTAAGGACA AGACCACTGG 840

CCAATCCAAG GGCTTCGCCT TCATCAGCTT CCACCGCCGC GAGGATGCTG CGCGTGCCAT 900

TGCCGGGGTG TCCGGCTTTG GCTACGACCA CCTCATCCTC AACGTCGAGT GGGCCAAGCC 960

GTCCACCAAC TAAGCCAGCT GCCACCGTGT ACTCGGTCCG GGACCCTTGG CGACAGAAGA 1020

CAGCCTCCGA GAGCGCGGGC TCCAAGGGCA ATAAAGCAGC TCCACTCTCA AAAAAAAAAA 1080

AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 1140

AAAAAAAAAA AA 1152

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 320 amino acids

(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

```
Met Pro Thr Gly Asp Phe Asp Ser Lys Pro Ser Trp Ala Asp Gln Val
1               5                  10                  15

Glu Glu Glu Gly Glu Asp Asp Lys Cys Val Thr Ser Glu Leu Leu Lys
            20                  25                  30

Gly Ile Pro Leu Ala Thr Gly Asp Thr Ser Pro Glu Pro Glu Leu Leu
        35                  40                  45

Pro Gly Ala Pro Leu Pro Pro Pro Lys Glu Val Ile Asn Gly Asn Ile
    50                  55                  60

Lys Thr Val Thr Glu Tyr Lys Ile Asp Glu Asp Gly Lys Lys Phe Lys
65                  70                  75                  80

Ile Val Arg Thr Phe Arg Ile Glu Thr Arg Lys Ala Ser Lys Ala Val
                85                  90                  95

Ala Arg Arg Lys Asn Trp Lys Lys Phe Gly Asn Ser Glu Phe Asp Pro
            100                 105                 110

Pro Gly Pro Asn Val Ala Thr Thr Thr Val Ser Asp Asp Val Ser Met
        115                 120                 125

Thr Phe Ile Thr Ser Lys Glu Asp Leu Asn Cys Gln Glu Glu Glu Asp
    130                 135                 140

Pro Met Asn Lys Leu Lys Gly Gln Lys Ile Val Ser Cys Arg Ile Cys
145                 150                 155                 160

Lys Gly Asp His Trp Thr Thr Arg Cys Pro Tyr Lys Asp Thr Leu Gly
                165                 170                 175

Pro Met Gln Lys Glu Leu Ala Glu Gln Leu Gly Leu Ser Thr Gly Glu
            180                 185                 190

Lys Glu Lys Leu Pro Gly Glu Leu Glu Pro Val Gln Ala Thr Gln Asn
        195                 200                 205

Lys Thr Gly Lys Tyr Val Pro Pro Ser Leu Arg Asp Gly Ala Ser Arg
    210                 215                 220

Arg Gly Glu Ser Met Gln Pro Thr Arg Arg Ala Asp Asp Asn Ala Thr
225                 230                 235                 240
```

```
Ile Arg Val Thr Asn Leu Ser Glu Asp Thr Arg Glu Thr Asp Leu Gln
                245                 250                 255

Glu Leu Phe Arg Pro Phe Gly Ser Ile Ser Arg Ile Tyr Leu Ala Lys
            260                 265                 270

Asp Lys Thr Thr Gly Gln Ser Lys Gly Phe Ala Phe Ile Ser Phe His
        275                 280                 285

Arg Arg Glu Asp Ala Ala Arg Ala Ile Ala Gly Val Ser Gly Phe Gly
    290                 295                 300

Tyr Asp His Leu Ile Leu Asn Val Glu Trp Ala Lys Pro Ser Thr Asn
305                 310                 315                 320
```

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1594 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

```
CTGAGACCTG GGCTGCTGTG AAAGCCCCTG CACAATCAGC CAGGGAGAAC TGGGCGGGTT     60

TAGTGGCCCC AGGCCCACTC CTCATGCAGC AGTGTGCTGG GGCGACAGCT CGTCTCCCCT    120

CTCTTAAGCA CCCGCTTCCT CACCACCCCC ACTGTTGGGC CTATAGTAGC AGGTTAGTGA    180

GTACCTAGGG CGGCTCAACT CCTCCCACAG CACCAACCCA GCATGGTCCC ACTGAAGTCC    240

TACTACGCCC TCCCCTCCCC AGCCTTTTCC AGAAACCATA CTGGGCTCAG ATCAGAGCTC    300

CGAAGCGGTC AAAGTGAGCT GAGCAGGACA GGCCCAGCCT TTCTCCACTG CCACGTCCCT    360

CATGCACATC ACTCATCTCC TGCTGCAGGC CAAGGCCAAA ATTGGGCTAG TCCTGGCCAG    420

GGAAATCAGA AGCTCTTCTT GGGTGAGATT GAGCCTCCTG TTGCTCCCTG GAGTTCCGGA    480

GGCTGGGCTG CAGCCCACTC AGCTTGCGGG CAAAATACGT GCTCTCCTCT CTCCTTGTCA    540

GCTGAGCAAA CCCAGGGAAT AGCCCTCCTC TCCCCAGGAA ACTTCTCTGA AATCTTAGAC    600

TTAGCCAGTC TTAGGCCTAC GATGCCACAC AAAGGTTGTT CAGGGAGAAG GGGGTGCAGG    660

AGGCAGAGGG TGCCCCGCAG GGAGCTGGTG GCTCCAGCCC CACTAGAGCT CCTAAAGATC    720
```

ACACAGCAGC TGCTCCTGAC AGGGATGCTC ATGCCCAGAA AGCAAGCCCA GGAGAGGAAG 780

GCAGAGTGTG ACAGAGCAGA GCCAGGGCCA GGCGCACCAG GAGAGGCGTT TCTGGGGCTC 840

CAGGGAAGTG CCACGGGAGG CAGAAGTCCA GAACTGCCCA TATAGATGCC CTTCTACATC 900

CTGGAGCCCA AATCAGTCAT GTGGGTGGGA AGTTCCCAGG GCAGTGGTCA CATCGTGAGA 960

ATTAGCAGGA AAGGCGGGGC CTTTCTTGTC ATAGCTATTT CTGAGGATGA AATGGGAGAC 1020

ATATGCCCAG CACCTGATGT AAGTTTATAT AATGTACCTA CCACTAAGAA ATACATGAAC 1080

CGTGCCATGA GGACAGTAAG TGTTCATAAA GCAACATGAA GCAAGAAACA GTGCAGGGTG 1140

CCCAGTGCAC ACACTAGAGA GAAATTGTGA ACATTAAGGA CAAGGAGAAT TGGTGTCTTT 1200

CTAAAACATA CTTATTTAAA AACACATACC CACTTACTAA TGTGGAATTA CACAGTTTGT 1260

AACAAGAAAA CAGTCTCTCC CATTCTCTAG TACTGYTCCC CTACCCAGCA GTCAMTTCCA 1320

GTTCATTCAG STATTTTTAA AATGTGCTTA TATGACTCTT GCTTGATATA TCAATYTTAG 1380

ACATTACCTG TTGACTCCCT GTTGTCATAC ATGAGGCTTT AGCTCTYTTT TGTCAGCAAC 1440

CCTCCCCCAT CCCTAGTTAT TAGGTTAAAA AATACTCAGA TTACTATTTC TATTACTATG 1500

TGAAAGTTAA CTGCGGAGCC AAGAGTTGGA CTATAATTAA ATTACCTTCC TTGTAAAAAA 1560

AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAA 1594

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 220 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

Met Val Pro Leu Lys Ser Tyr Tyr Ala Leu Pro Ser Pro Ala Phe Ser
1 5 10 15

Arg Asn His Thr Gly Leu Arg Ser Glu Leu Arg Ser Gly Gln Ser Glu
20 25 30

Leu Ser Arg Thr Gly Pro Ala Phe Leu His Cys His Val Pro His Ala
35 40 45

His His Ser Ser Pro Ala Ala Gly Gln Gly Gln Asn Trp Ala Ser Pro

```
     50                  55                  60

Gly Gln Gly Asn Gln Lys Leu Phe Leu Gly Glu Ile Glu Pro Pro Val
65                  70                  75                  80

Ala Pro Trp Ser Ser Gly Gly Trp Ala Ala Ala His Ser Ala Cys Gly
                85                  90                  95

Gln Asn Thr Cys Ser Pro Leu Ser Leu Ser Ala Glu Gln Thr Gln Gly
            100                 105                 110

Ile Ala Leu Leu Ser Pro Gly Asn Phe Ser Glu Ile Leu Asp Leu Ala
        115                 120                 125

Ser Leu Arg Pro Thr Met Pro His Lys Gly Cys Ser Gly Arg Arg Gly
    130                 135                 140

Cys Arg Arg Gln Arg Val Pro Arg Arg Glu Leu Val Ala Pro Ala Pro
145                 150                 155                 160

Leu Glu Leu Leu Lys Ile Thr Gln Gln Leu Leu Leu Thr Gly Met Leu
                165                 170                 175

Met Pro Arg Lys Gln Ala Gln Glu Arg Lys Ala Glu Cys Asp Arg Ala
            180                 185                 190

Glu Pro Gly Pro Gly Ala Pro Gly Glu Ala Phe Leu Gly Leu Gln Gly
        195                 200                 205

Ser Ala Thr Gly Gly Arg Ser Pro Glu Leu Pro Ile
    210                 215                 220
```

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "oligonucleotide"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

TNAATAAACTG GACGGATGCA CTGATAGG 29

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "oligonucleotide"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

CNCTGATAACA AAGCATTGCC ACTGGCGC 29

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "oligonucleotide"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

TNATCCAGAAA ATTACCGCCG TCCGACCG 29

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "oligonucleotide"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

CNCTTAGAAGC CTTCATTTTG GGAAGTGC 29

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "oligonucleotide"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

CNGAGAAGACT CAACGAGGCA GCCAAGAA 29

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "oligonucleotide"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

CNTGCTGACTT GGCCCAAGAA GCTTGATT 29

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "oligonucleotide"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

GNGCTGCTTTC CAGACTCCTT CAGTTTCT 29

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "oligonucleotide"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

ANCCACAGCGT GGTTCTTGAG GTGTTCCC 29

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "oligonulceotide"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

GNTCTTCTGGC CCTTGAGTTT GTTCATAG 29

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "oligonucleotide"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

GNTGAGCCGCC CTAGGTACTC ACTAACCT 29

What is claimed is:

1. An isolated polynucleotide selected from the group consisting of:

(a) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:1;

(b) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:1 from nucleotide 1799 to nucleotide 2332;

(c) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:1 from nucleotide 2288 to nucleotide 2332;

(d) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:1 from nucleotide 2306 to nucleotide 2754;

(e) a polynucleotide comprising the nucleotide sequence of the full-length protein coding sequence of clone en539_8 deposited under accession number ATCC 98408;

(f) a polynucleotide encoding the full-length protein encoded by the cDNA insert of clone en539_8 deposited under accession number ATCC 98408;

(g) a polynucleotide comprising the nucleotide sequence of a mature protein coding sequence of clone en539_8 deposited under accession number ATCC 98408;

(h) a polynucleotide encoding a mature protein encoded by the cDNA insert of clone en539_8 deposited under accession number ATCC 98408;

(i) a polynucleotide encoding a protein comprising the amino acid sequence of SEQ ID NO:2;

(j) a polynucleotide encoding a protein comprising a fragment of the amino acid sequence of SEQ ID NO:2 having biological activity, the fragment comprising the amino acid sequence from amino acid 84 to amino acid 93 of SEQ ID NO:2;

(k) a polynucleotide which is an allelic variant of a polynucleotide of (a)-(h) above;

(l) a polynucleotide which encodes a species homologue of the protein of (i) or (j) above ; and

(m) a polynucleotide that hybridizes under stringent conditions to any one of the polynucleotides specified in (a)-(j).

2. The polynucleotide of claim 1 wherein said polynucleotide is operably linked to at least one expression control sequence.

3. A host cell transformed with the polynucleotide of claim 2.

4. The host cell of claim 3, wherein said cell is a mammalian cell.

5. A process for producing a protein encoded by the polynucleotide of claim 2, which process comprises:

(a) growing a culture of the host cell of claim 3 in a suitable culture medium; and

(b) purifying said protein from the culture.

6. A protein produced according to the process of claim 5.

7. The protein of claim 6 comprising a mature protein.

8. A protein comprising an amino acid sequence selected from the group consisting of:

(a) the amino acid sequence of SEQ ID NO:2;

(b) the amino acid sequence of SEQ ID NO:2 from amino acid 169 to amino acid 178;

(c) fragments of the amino acid sequence of SEQ ID NO:2 comprising the amino acid sequence from amino acid 84 to amino acid 93 of SEQ ID NO:2; and

(d) the amino acid sequence encoded by the cDNA insert of clone en539_8 deposited under accession number ATCC 98408;

the protein being substantially free from other mammalian proteins.

9. The protein of claim 8, wherein said protein comprises the amino acid sequence of SEQ ID NO:2.

10. The protein of claim 8, wherein said protein comprises the amino acid sequence of SEQ ID NO:2 from amino acid 169 to amino acid 178.

11. A composition comprising the protein of claim 8 and a pharmaceutically acceptable carrier.

12. An isolated gene corresponding to the cDNA sequence of SEQ ID NO:1.

13. An isolated polynucleotide selected from the group consisting of:

(a) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:3;

(b) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:3 from nucleotide 91 to nucleotide 966;

(c) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:3 from nucleotide 1 to nucleotide 337;

(d) a polynucleotide comprising the nucleotide sequence of the full-length protein coding sequence of clone eq188_1 deposited under accession number ATCC 98408;

(e) a polynucleotide encoding the full-length protein encoded by the cDNA insert of clone eq188_1 deposited under accession number ATCC 98408;

(f) a polynucleotide comprising the nucleotide sequence of a mature protein coding sequence of clone eq188_1 deposited under accession number ATCC 98408;

(g) a polynucleotide encoding a mature protein encoded by the cDNA insert of clone eq188_1 deposited under accession number ATCC 98408;

(h) a polynucleotide encoding a protein comprising the amino acid sequence of SEQ ID NO:4;

(i) a polynucleotide encoding a protein comprising a fragment of the amino acid sequence of SEQ ID NO:4 having biological activity, the fragment comprising the amino acid sequence from amino acid 141 to amino acid 150 of SEQ ID NO:4;

(j) a polynucleotide which is an allelic variant of a polynucleotide of (a)-(g) above;

(k) a polynucleotide which encodes a species homologue of the protein of (h) or (i) above ; and

(l) a polynucleotide that hybridizes under stringent conditions to any one of the polynucleotides specified in (a)-(i).

14. A protein comprising an amino acid sequence selected from the group consisting of:

(a) the amino acid sequence of SEQ ID NO:4;

(b) the amino acid sequence of SEQ ID NO:4 from amino acid 1 to amino acid 83;

(c) fragments of the amino acid sequence of SEQ ID NO:4 comprising the amino acid sequence from amino acid 141 to amino acid 150 of SEQ ID NO:4; and

(d) the amino acid sequence encoded by the cDNA insert of clone eq188_1 deposited under accession number ATCC 98408;

the protein being substantially free from other mammalian proteins.

15. An isolated gene corresponding to the cDNA sequence of SEQ ID NO:3.

16. An isolated polynucleotide selected from the group consisting of:

(a) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:5;

(b) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:5 from nucleotide 51 to nucleotide 1358;

(c) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:5 from nucleotide 99 to nucleotide 1358;

(d) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:5 from nucleotide 249 to nucleotide 566;

(e) a polynucleotide comprising the nucleotide sequence of the full-length protein coding sequence of clone er80_1 deposited under accession number ATCC 98408;

(f) a polynucleotide encoding the full-length protein encoded by the cDNA insert of clone er80_1 deposited under accession number ATCC 98408;

(g) a polynucleotide comprising the nucleotide sequence of a mature protein coding sequence of clone er80_1 deposited under accession number ATCC 98408;

(h) a polynucleotide encoding a mature protein encoded by the cDNA insert of clone er80_1 deposited under accession number ATCC 98408;

(i) a polynucleotide encoding a protein comprising the amino acid sequence of SEQ ID NO:6;

(j) a polynucleotide encoding a protein comprising a fragment of the amino acid sequence of SEQ ID NO:6 having biological activity, the fragment comprising the amino acid sequence from amino acid 213 to amino acid 222 of SEQ ID NO:6;

(k) a polynucleotide which is an allelic variant of a polynucleotide of (a)-(h) above;

(l) a polynucleotide which encodes a species homologue of the protein of (i) or (j) above ; and

(m) a polynucleotide that hybridizes under stringent conditions to any one of the polynucleotides specified in (a)-(j).

17. A protein comprising an amino acid sequence selected from the group consisting of:

(a) the amino acid sequence of SEQ ID NO:6;

(b) the amino acid sequence of SEQ ID NO:6 from amino acid 1 to amino acid 172;

(c) fragments of the amino acid sequence of SEQ ID NO:6 comprising the amino acid sequence from amino acid 213 to amino acid 222 of SEQ ID NO:6; and

(d) the amino acid sequence encoded by the cDNA insert of clone er80_1 deposited under accession number ATCC 98408;

the protein being substantially free from other mammalian proteins.

18. An isolated gene corresponding to the cDNA sequence of SEQ ID NO:5.

19. An isolated polynucleotide selected from the group consisting of:

(a) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:7;

(b) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:7 from nucleotide 571 to nucleotide 3306;

(c) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:7 from nucleotide 726 to nucleotide 1320;

(d) a polynucleotide comprising the nucleotide sequence of the full-length protein coding sequence of clone er418_5 deposited under accession number ATCC 98408;

(e) a polynucleotide encoding the full-length protein encoded by the cDNA insert of clone er418_5 deposited under accession number ATCC 98408;

(f) a polynucleotide comprising the nucleotide sequence of a mature protein coding sequence of clone er418_5 deposited under accession number ATCC 98408;

(g) a polynucleotide encoding a mature protein encoded by the cDNA insert of clone er418_5 deposited under accession number ATCC 98408;

(h) a polynucleotide encoding a protein comprising the amino acid sequence of SEQ ID NO:8;

(i) a polynucleotide encoding a protein comprising a fragment of the amino acid sequence of SEQ ID NO:8 having biological activity, the fragment comprising the amino acid sequence from amino acid 450 to amino acid 459 of SEQ ID NO:8;

(j) a polynucleotide which is an allelic variant of a polynucleotide of (a)-(g) above;

(k) a polynucleotide which encodes a species homologue of the protein of (h) or (i) above ; and

(l) a polynucleotide that hybridizes under stringent conditions to any one of the polynucleotides specified in (a)-(i).

20. A protein comprising an amino acid sequence selected from the group consisting of:

(a) the amino acid sequence of SEQ ID NO:8;

(b) the amino acid sequence of SEQ ID NO:8 from amino acid 71 to amino acid 250;

(c) fragments of the amino acid sequence of SEQ ID NO:8 comprising the amino acid sequence from amino acid 450 to amino acid 459 of SEQ ID NO:8; and

(d) the amino acid sequence encoded by the cDNA insert of clone er418_5 deposited under accession number ATCC 98408;

the protein being substantially free from other mammalian proteins.

21. An isolated gene corresponding to the cDNA sequence of SEQ ID NO:7.

22. An isolated polynucleotide selected from the group consisting of:

(a) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:9;

(b) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:9 from nucleotide 503 to nucleotide 2770;

(c) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:9 from nucleotide 572 to nucleotide 2770;

(d) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:9 from nucleotide 490 to nucleotide 772;

(e) a polynucleotide comprising the nucleotide sequence of the full-length protein coding sequence of clone fa252_8 deposited under accession number ATCC 98408;

(f) a polynucleotide encoding the full-length protein encoded by the cDNA insert of clone fa252_8 deposited under accession number ATCC 98408;

(g) a polynucleotide comprising the nucleotide sequence of a mature protein coding sequence of clone fa252_8 deposited under accession number ATCC 98408;

(h) a polynucleotide encoding a mature protein encoded by the cDNA insert of clone fa252_8 deposited under accession number ATCC 98408;

(i) a polynucleotide encoding a protein comprising the amino acid sequence of SEQ ID NO:10;

(j) a polynucleotide encoding a protein comprising a fragment of the amino acid sequence of SEQ ID NO:10 having biological activity, the fragment comprising the amino acid sequence from amino acid 373 to amino acid 382 of SEQ ID NO:10;

(k) a polynucleotide which is an allelic variant of a polynucleotide of (a)-(h) above;

(l) a polynucleotide which encodes a species homologue of the protein of (i) or (j) above ; and

(m) a polynucleotide that hybridizes under stringent conditions to any one of the polynucleotides specified in (a)-(j).

23. A protein comprising an amino acid sequence selected from the group consisting of:

(a) the amino acid sequence of SEQ ID NO:10;

(b) the amino acid sequence of SEQ ID NO:10 from amino acid 1 to amino acid 90;

(c) fragments of the amino acid sequence of SEQ ID NO:10 comprising the amino acid sequence from amino acid 373 to amino acid 382 of SEQ ID NO:10; and

(d) the amino acid sequence encoded by the cDNA insert of clone fa252_8 deposited under accession number ATCC 98408;

the protein being substantially free from other mammalian proteins.

24. An isolated gene corresponding to the cDNA sequence of SEQ ID NO:9.

25. An isolated polynucleotide selected from the group consisting of:

(a) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:11;

(b) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:11 from nucleotide 104 to nucleotide 565;

(c) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:11 from nucleotide 1 to nucleotide 501;

(d) a polynucleotide comprising the nucleotide sequence of the full-length protein coding sequence of clone fg912_1 deposited under accession number ATCC 98408;

(e) a polynucleotide encoding the full-length protein encoded by the cDNA insert of clone fg912_1 deposited under accession number ATCC 98408;

(f) a polynucleotide comprising the nucleotide sequence of a mature protein coding sequence of clone fg912_1 deposited under accession number ATCC 98408;

(g) a polynucleotide encoding a mature protein encoded by the cDNA insert of clone fg912_1 deposited under accession number ATCC 98408;

(h) a polynucleotide encoding a protein comprising the amino acid sequence of SEQ ID NO:12;

(i) a polynucleotide encoding a protein comprising a fragment of the amino acid sequence of SEQ ID NO:12 having biological activity, the fragment comprising the amino acid sequence from amino acid 72 to amino acid 81 of SEQ ID NO:12;

(j) a polynucleotide which is an allelic variant of a polynucleotide of (a)-(g) above;

(k) a polynucleotide which encodes a species homologue of the protein of (h) or (i) above ; and

(l) a polynucleotide that hybridizes under stringent conditions to any one of the polynucleotides specified in (a)-(i).

26. A protein comprising an amino acid sequence selected from the group consisting of:

(a) the amino acid sequence of SEQ ID NO:12;

(b) the amino acid sequence of SEQ ID NO:12 from amino acid 1 to amino acid 132;

(c) fragments of the amino acid sequence of SEQ ID NO:12 comprising the amino acid sequence from amino acid 72 to amino acid 81 of SEQ ID NO:12; and

(d) the amino acid sequence encoded by the cDNA insert of clone fg912_1 deposited under accession number ATCC 98408;

the protein being substantially free from other mammalian proteins.

27. An isolated gene corresponding to the cDNA sequence of SEQ ID NO:11.

28. An isolated polynucleotide selected from the group consisting of:

(a) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:13;

(b) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:13 from nucleotide 77 to nucleotide 1093;

(c) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:13 from nucleotide 167 to nucleotide 1093;

(d) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:13 from nucleotide 1 to nucleotide 718;

(e) a polynucleotide comprising the nucleotide sequence of the full-length protein coding sequence of clone fg949_3 deposited under accession number ATCC 98408;

(f) a polynucleotide encoding the full-length protein encoded by the cDNA insert of clone fg949_3 deposited under accession number ATCC 98408;

(g) a polynucleotide comprising the nucleotide sequence of a mature protein coding sequence of clone fg949_3 deposited under accession number ATCC 98408;

(h) a polynucleotide encoding a mature protein encoded by the cDNA insert of clone fg949_3 deposited under accession number ATCC 98408;

(i) a polynucleotide encoding a protein comprising the amino acid sequence of SEQ ID NO:14;

(j) a polynucleotide encoding a protein comprising a fragment of the amino acid sequence of SEQ ID NO:14 having biological activity, the fragment comprising the amino acid sequence from amino acid 164 to amino acid 173 of SEQ ID NO:14;

(k) a polynucleotide which is an allelic variant of a polynucleotide of (a)-(h) above;

(l) a polynucleotide which encodes a species homologue of the protein of (i) or (j) above ; and

(m) a polynucleotide that hybridizes under stringent conditions to any one of the polynucleotides specified in (a)-(j).

29. A protein comprising an amino acid sequence selected from the group consisting of:

(a) the amino acid sequence of SEQ ID NO:14;

(b) the amino acid sequence of SEQ ID NO:14 from amino acid 1 to amino acid 214;

(c) fragments of the amino acid sequence of SEQ ID NO:14 comprising the amino acid sequence from amino acid 164 to amino acid 173 of SEQ ID NO:14; and

(d) the amino acid sequence encoded by the cDNA insert of clone fg949_3 deposited under accession number ATCC 98408;

the protein being substantially free from other mammalian proteins.

30. An isolated gene corresponding to the cDNA sequence of SEQ ID NO:13.

31. An isolated polynucleotide selected from the group consisting of:

(a) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:15;

(b) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:15 from nucleotide 19 to nucleotide 1023;

(c) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:15 from nucleotide 247 to nucleotide 711;

(d) a polynucleotide comprising the nucleotide sequence of the full-length protein coding sequence of clone fk354_4 deposited under accession number ATCC 98408;

(e) a polynucleotide encoding the full-length protein encoded by the cDNA insert of clone fk354_4 deposited under accession number ATCC 98408;

(f) a polynucleotide comprising the nucleotide sequence of a mature protein coding sequence of clone fk354_4 deposited under accession number ATCC 98408;

(g) a polynucleotide encoding a mature protein encoded by the cDNA insert of clone fk354_4 deposited under accession number ATCC 98408;

(h) a polynucleotide encoding a protein comprising the amino acid sequence of SEQ ID NO:16;

(i) a polynucleotide encoding a protein comprising a fragment of the amino acid sequence of SEQ ID NO:16 having biological activity, the fragment comprising the amino acid sequence from amino acid 162 to amino acid 171 of SEQ ID NO:16;

(j) a polynucleotide which is an allelic variant of a polynucleotide of (a)-(g) above;

(k) a polynucleotide which encodes a species homologue of the protein of (h) or (i) above ; and

(l) a polynucleotide that hybridizes under stringent conditions to any one of the polynucleotides specified in (a)-(i).

32. A protein comprising an amino acid sequence selected from the group consisting of:

(a) the amino acid sequence of SEQ ID NO:16;

(b) the amino acid sequence of SEQ ID NO:16 from amino acid 147 to amino acid 231;

(c) fragments of the amino acid sequence of SEQ ID NO:16 comprising the amino acid sequence from amino acid 162 to amino acid 171 of SEQ ID NO:16; and

(d) the amino acid sequence encoded by the cDNA insert of clone fk354_4 deposited under accession number ATCC 98408;

the protein being substantially free from other mammalian proteins.

33. An isolated gene corresponding to the cDNA sequence of SEQ ID NO:15.

34. An isolated polynucleotide selected from the group consisting of:

(a) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:17;

(b) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:17 from nucleotide 11 to nucleotide 970;

(c) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:17 from nucleotide 1 to nucleotide 575;

(d) a polynucleotide comprising the nucleotide sequence of the full-length protein coding sequence of clone fm150_1 deposited under accession number ATCC 98408;

(e) a polynucleotide encoding the full-length protein encoded by the cDNA insert of clone fm150_1 deposited under accession number ATCC 98408;

(f) a polynucleotide comprising the nucleotide sequence of a mature protein coding sequence of clone fm150_1 deposited under accession number ATCC 98408;

(g) a polynucleotide encoding a mature protein encoded by the cDNA insert of clone fm150_1 deposited under accession number ATCC 98408;

(h) a polynucleotide encoding a protein comprising the amino acid sequence of SEQ ID NO:18;

(i) a polynucleotide encoding a protein comprising a fragment of the amino acid sequence of SEQ ID NO:18 having biological activity, the fragment

comprising the amino acid sequence from amino acid 155 to amino acid 164 of SEQ ID NO:18;

(j) a polynucleotide which is an allelic variant of a polynucleotide of (a)-(g) above;

(k) a polynucleotide which encodes a species homologue of the protein of (h) or (i) above ; and

(l) a polynucleotide that hybridizes under stringent conditions to any one of the polynucleotides specified in (a)-(i).

35. A protein comprising an amino acid sequence selected from the group consisting of:

(a) the amino acid sequence of SEQ ID NO:18;

(b) the amino acid sequence of SEQ ID NO:18 from amino acid 1 to amino acid 188;

(c) fragments of the amino acid sequence of SEQ ID NO:18 comprising the amino acid sequence from amino acid 155 to amino acid 164 of SEQ ID NO:18; and

(d) the amino acid sequence encoded by the cDNA insert of clone fm150_1 deposited under accession number ATCC 98408;

the protein being substantially free from other mammalian proteins.

36. An isolated gene corresponding to the cDNA sequence of SEQ ID NO:17.

37. An isolated polynucleotide selected from the group consisting of:

(a) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:19;

(b) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:19 from nucleotide 223 to nucleotide 882;

(c) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:19 from nucleotide 46 to nucleotide 351;

(d) a polynucleotide comprising the nucleotide sequence of the full-length protein coding sequence of clone gu534_1 deposited under accession number ATCC 98408;

(e) a polynucleotide encoding the full-length protein encoded by the cDNA insert of clone gu534_1 deposited under accession number ATCC 98408;

(f) a polynucleotide comprising the nucleotide sequence of a mature protein coding sequence of clone gu534_1 deposited under accession number ATCC 98408;

(g) a polynucleotide encoding a mature protein encoded by the cDNA insert of clone gu534_1 deposited under accession number ATCC 98408;

(h) a polynucleotide encoding a protein comprising the amino acid sequence of SEQ ID NO:20;

(i) a polynucleotide encoding a protein comprising a fragment of the amino acid sequence of SEQ ID NO:20 having biological activity, the fragment comprising the amino acid sequence from amino acid 105 to amino acid 114 of SEQ ID NO:20;

(j) a polynucleotide which is an allelic variant of a polynucleotide of (a)-(g) above;

(k) a polynucleotide which encodes a species homologue of the protein of (h) or (i) above ; and

(l) a polynucleotide that hybridizes under stringent conditions to any one of the polynucleotides specified in (a)-(i).

38. A protein comprising an amino acid sequence selected from the group consisting of:

(a) the amino acid sequence of SEQ ID NO:20;

(b) the amino acid sequence of SEQ ID NO:20 from amino acid 1 to amino acid 43;

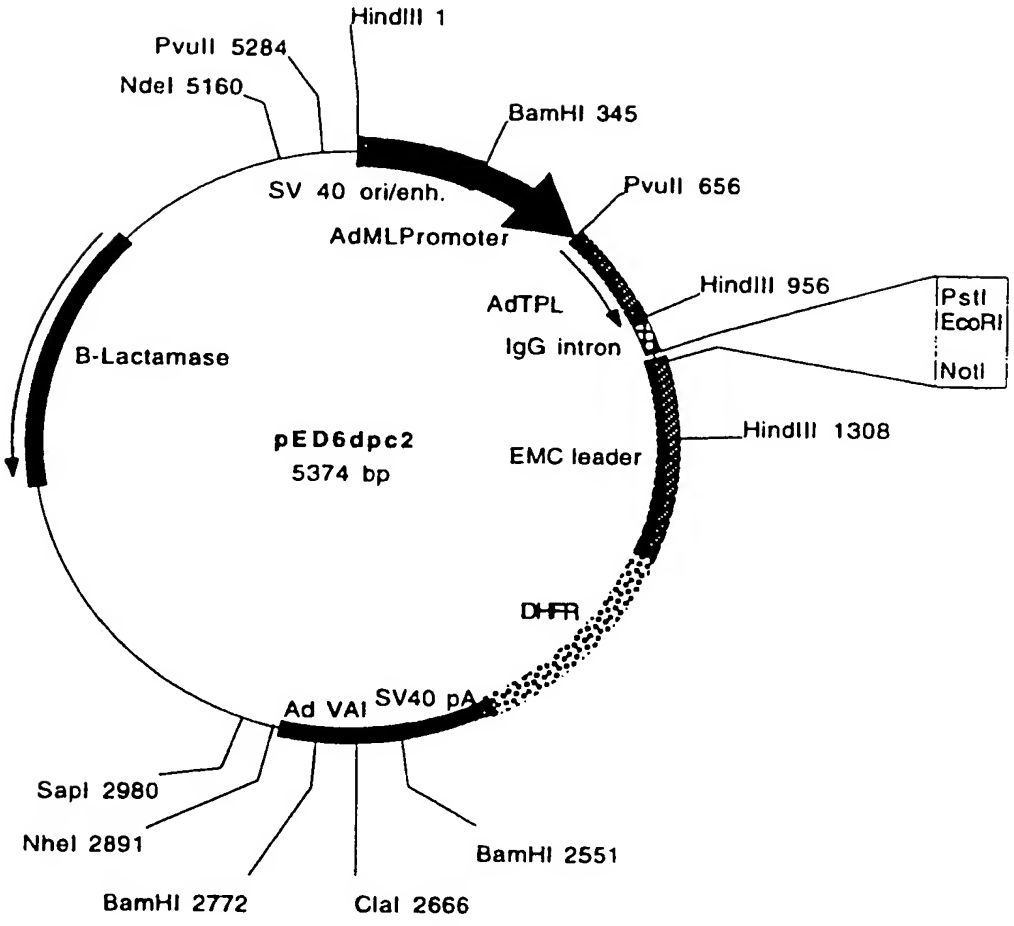
(c) fragments of the amino acid sequence of SEQ ID NO:20 comprising the amino acid sequence from amino acid 105 to amino acid 114 of SEQ ID NO:20; and

(d) the amino acid sequence encoded by the cDNA insert of clone gu534_1 deposited under accession number ATCC 98408;

the protein being substantially free from other mammalian proteins.

39. An isolated gene corresponding to the cDNA sequence of SEQ ID NO:19.

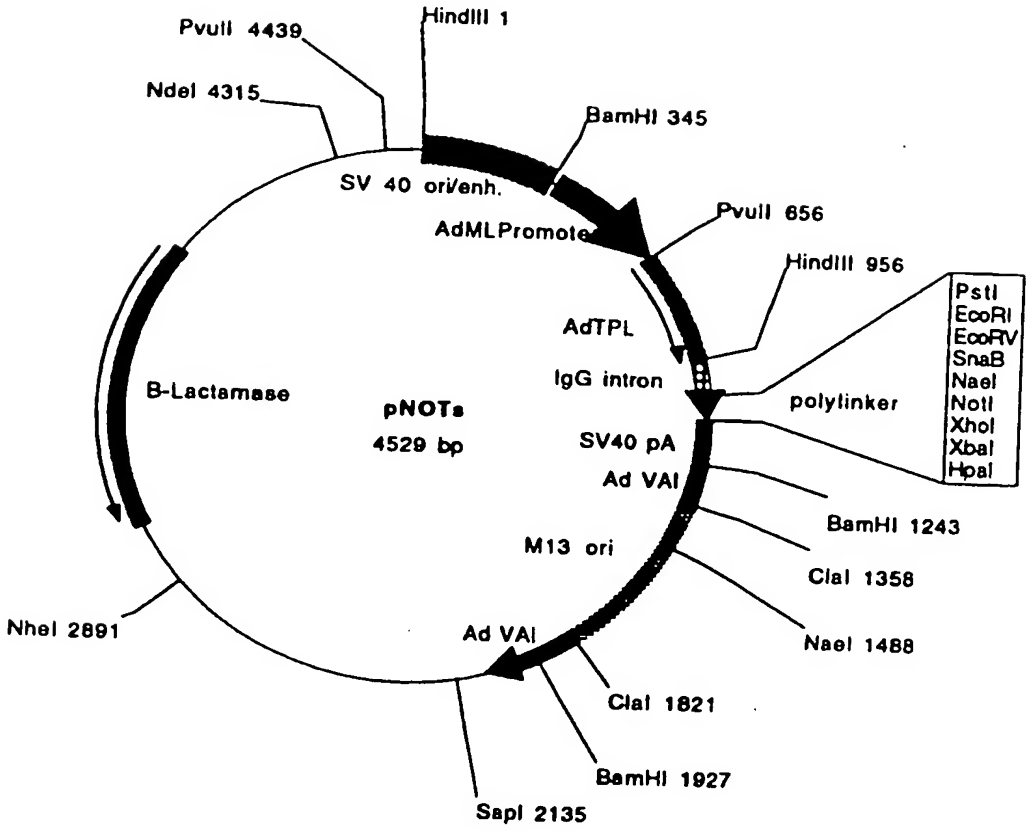
FIGURE 1A



Plasmid name: pED6dpc2
Plasmid size: 5374 bp

Comments/References: pED6dpc2 is derived from pED6dpc1 by insertion of a new polylinker to facilitate cDNA cloning. SST cDNAs are cloned between EcoRI and NotI. pED vectors are described in Kaufman et al.(1991), NAR 19: 4485-4490.

FIGURE 1B



Plasmid name: pNOTs
Plasmid size: 4529 bp

Comments/References: pNOTs is a derivative of pMT2 (Kaufman et al,1989. Mol.Cell.Biol.9:1741-1750). DHFR was deleted and a new polylinker was inserted between EcoRI and HpaI. M13 origin of replication was inserted in the ClaI site. SST cDNAs are cloned between EcoRI and NotI

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/12, C07K 14/47, A61K 38/17

A3

(11) International Publication Number: WO 98/46757

(43) International Publication Date: 22 October 1998 (22.10.98)

(21) International Application Number: PCT/US98/07999

(22) International Filing Date: 14 April 1998 (14.04.98)

(30) Priority Data:

08/843,374	15 April 1997 (15.04.97)	US
09/059,487	13 April 1998 (13.04.98)	US

(71) Applicant: GENETICS INSTITUTE, INC. [US/US]; 87 CambridgePark Drive, Cambridge, MA 02140 (US).

(72) Inventors: JACOBS, Kenneth; 151 Beaumont Avenue, Newton, MA 02160 (US). MCCOY, John, M.; 56 Howard Street, Reading, MA 01867 (US). LAVALLIE, Edward, R.; 113 Ann Lee Road, Harvard, MA 01451 (US). RACIE, Lisa, A.; 124 School Street, Acton, MA 01720 (US). MERBERG, David; 2 Orchard Drive, Acton, MA 01720 (US). TREACY, Maurice; 93 Walcott Road, Chestnut Hill, MA 02167 (US). SPAULDING, Vikki; 11 Meadowbank Road, Billeria, MA 01821 (US). AGOSTINO, Michael, J.; 26 Wolcott Avenue, Andover, MA 01810 (US).

(74) Agent: SPRUNGER, Suzanne, A.; Genetics Institute, Inc., 87 CambridgePark Drive, Cambridge, MA 02140 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(88) Date of publication of the international search report:
18 February 1999 (18.02.99)

(54) Title: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM

(57) Abstract

Polynucleotides and the proteins encoded thereby are disclosed.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

INTERNATIONAL SEARCH REPORT

Inter nal Application No
PCT/US 98/07999

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/12 C07K14/47 A61K38/17

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12N C07K A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Database EMBL EMEST4, Entry HS08424, Accession number T54084 28 February 1995 85% identity with Seq.ID:1 nt.19-386 XP002070766 cited in the application see the whole document ---	1,12
A	WO 97 07198 A (GENETICS INSTITUTE INC.) 27 February 1997 ---	
A	US 5 536 637 A (JACOBS KENNETH) 16 July 1996 cited in the application -----	

☐ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
8 July 1998	09. 10. 98

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Macchia, G

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

I ıational application No.
PCT/US 98/ 07999

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see further information sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1-12

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: 1-12

   Polynucleotide comprising the nucleotide sequence of Seq.ID:1 and encoding a polypeptide of Seq.ID:2 or fragments. Polynucleotide fragments, variants, homologues, gene thereof. Host cell transformed with said polynucleotide. Protein comprising an amino acid sequence of Seq.ID:2 or fragments thereof. Process for producing said protein.

2. Claims: 13-15

   Polynucleotide comprising the nucleotide sequence of Seq.ID:3 and encoding a polypeptide of Seq.ID:4 or fragments. Polynucleotide fragments, variants, homologues, gene thereof. Protein comprising an amino acid sequence of Seq.ID:4 or fragments thereof.

3. Claims: 16-18

   As invention 2 but concerning Seq.ID:5 and 6.

4. Claims: 19-21

   As invention 2 but concerning Seq.ID:7 and 8.

5. Claims: 22-24

   As invention 2 but concerning Seq.ID:9 and 10.

6. Claims: 25-27

   As invention 2 but concerning Seq.ID:11 and 12.

7. Claims: 28-30

   As invention 2 but concerning Seq.ID:13 and 14.

8. Claims: 31-33

   As invention 2 but concerning Seq.ID:15 and 16.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

9. Claims: 34-36

As invention 2 but concerning Seq.ID:17 and 18.

10. Claims: 37-39

As invention 2 but concerning Seq.ID:19 and 20.

INTERNATIONAL SEARCH REPORT

mation on patent family members

Interr nal Application No
PCT/US 98/07999

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9707198 A	27-02-97	US 5707829 A	13-01-98
		AU 6712396 A	18-02-97
		AU 6768596 A	12-03-97
		EP 0839196 A	06-05-98
		EP 0851875 A	08-07-98
		WO 9704097 A	06-02-97
US 5536637 A	16-07-96	US 5712116 A	27-01-98